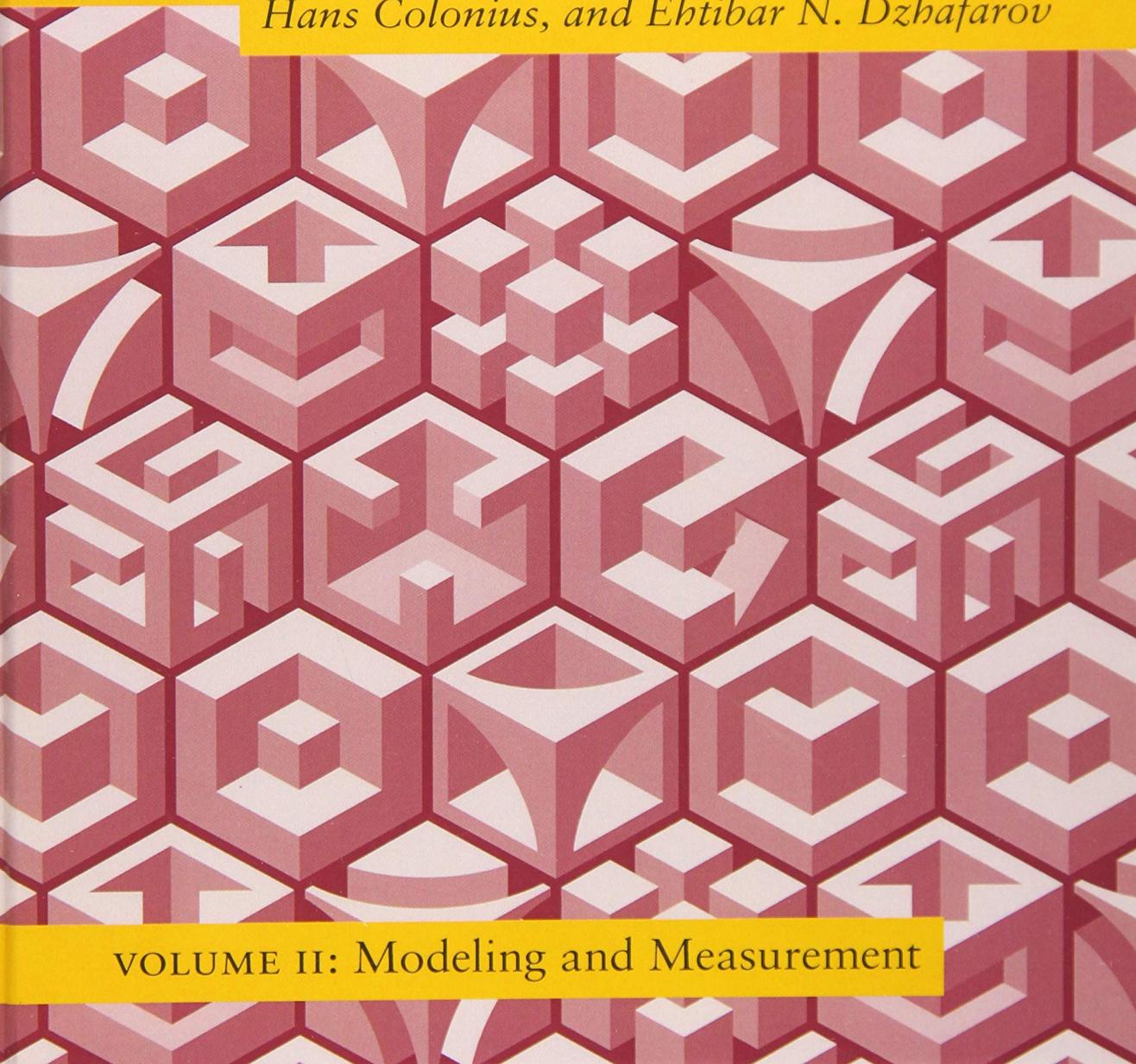


NEW HANDBOOK OF
MATHEMATICAL
PSYCHOLOGY

*Edited by William H. Batchelder,
Hans Colonius, and Ehtibar N. Dzhafarov*



VOLUME II: Modeling and Measurement

New Handbook of Mathematical Psychology

Volume 2. Modeling and Measurement

The field of mathematical psychology began in the 1950s and includes both psychological theorizing, in which mathematics plays a key role, and applied mathematics motivated by substantive problems in psychology. Central to its success was the publication of the first *Handbook of Mathematical Psychology* in the 1960s. The psychological sciences have since expanded to include new areas of research, and significant advances have been made in both traditional psychological domains and in the applications of the computational sciences to psychology. Upholding the rigor of the original handbook, the *New Handbook of Mathematical Psychology* reflects the current state of the field by exploring the mathematical and computational foundations of new developments over the last half-century. The second volume focuses on areas of mathematics that are used in constructing models of cognitive phenomena and decision-making, and on the role of measurement in psychology.

WILLIAM H. BATCHELDER is Professor of Cognitive Sciences at the University of California, Irvine.

HANS COLONIUS is Professor of Psychology at Oldenburg University, Germany.

EHTIBAR N. DZHAFAROV is Professor of Psychological Sciences at Purdue University.

New Handbook of Mathematical Psychology

Volume 2. Modeling and Measurement

Edited by

William H. Batchelder

Hans Colonius

Ehtibar N. Dzhafarov



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107029071

DOI: 10.1017/9781139245906

© Cambridge University Press 2018

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2018

Printed in the United Kingdom by TJ International Ltd, Padstow, Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Control Number: 2017303272

ISBN 978-1-107-02907-1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of
URLs for external or third-party internet websites referred to in this publication
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

Contents

<i>List of contributors</i>	<i>page</i> vi
<i>List of abbreviations</i>	vii
<i>Preface</i>	ix
WILLIAM H. BATCHELDER, HANS COLONIUS, AND EHTIBAR N. DZHAFAROV	
1 Stochastic Methods for Modeling Decision-making	1
ADELE DIEDERICH AND KEIVAN MALLAHI-KARAI	
2 The Diffusion Model of Speeded Choice, from a Rational Perspective	71
MATT JONES	
3 Stochastic Foundations of Elementary Mental Architectures	104
JOSEPH W. HOUPT, JAMES T. TOWNSEND, AND BRETT JEFFERSON	
4 Identifiability of Probabilistic Models, with Examples from Knowledge Structure Theory	128
JEAN-PAUL DOIGNON, JÜRGEN HELLER, AND LUCA STEFANUTTI	
5 Quantum Models of Cognition and Decision	185
JEROME R. BUSEMEYER AND PETER D. KVAM	
6 Computational Cognitive Neuroscience	223
F. GREGORY ASHBY	
7 Discovering Aggregation Properties via Voting	271
DONALD G. SAARI	
8 Categorization Based on Similarity and Features: The Reproducing Kernel Banach Space (RKBS) Approach	322
JUN ZHANG AND HAIZHANG ZHANG	
9 The Axiom of Meaningfulness in Science and Geometry	374
JEAN-CLAUDE FALMAGNE, LOUIS NARENS, AND CHRISTOPHER DOBLE	
<i>Index</i>	457

Contributors

F. GREGORY ASHBY, University of California, Santa Barbara (USA)
WILLIAM H. BATCHELDER, University of California, Irvine (USA)
JEROME R. BUSEMEYER, Indiana University, Bloomington (USA)
HANS COLONIUS, Oldenburg University (Germany)
ADELE DIEDERICH, Jacobs University Bremen (Germany)
CHRISTOPHER DOBLE, McGraw-Hill Education (USA)
JEAN-PAUL DOIGNON, Université Libre de Bruxelles (Belgium)
EHTIBAR N. DZHAFAROV, Purdue University (USA)
JEAN-CLAUDE FALMAGNE, University of California, Irvine (USA)
JOSEPH W. HOUPP, Wright State University (USA)
JÜRGEN HELLER, University of Tübingen (Germany)
BRETT JEFFERSON, Indiana University, Bloomington (USA)
MATT JONES, University of Colorado, Boulder (USA)
PETER D. KVAM, Michigan State University (USA)
KEIVAN MALLAHI-KARAI, Jacobs University Bremen (Germany)
LOUIS NARENS, University of California, Irvine (USA)
DONALD G. SAARI, University of California, Irvine (USA)
LUCA STEFANUTTI, University of Padova (Italy)
JAMES T. TOWNSEND, Indiana University, Bloomington (USA)
HAIZHANG ZHANG, Sun Yat-Sen University, Guangzhou (China)
JUN ZHANG, University of Michigan (USA)

Abbreviations

- AHP, analytic hierarchy process
BAR, Borda assignment rule
BLIM, basic local independence model
CCN, computational cognitive neuroscience
cdf, cumulative distribution function
CRM, correct response model
DA, dopamine
DAT, DA active transporter
fMRI, functional magnetic resonance imaging
GLM, general linear model
Hrf, hemodynamic response function
IIA, independence of irrelevant alternatives
KL, Kullback–Leibler
KST, knowledge structure theory
LATER, linear approach to threshold with ergodic rate
LBA, linear ballistic accumulator
LCA, leaky competing accumulator
LCM, latent class model
LTD, long-term depression
LTP, long-term potentiation
MAP, maximum *a posteriori*
OU, Ornstein–Uhlenbeck
PCC, principle of correspondent change
pdf, probability (mass) distribution function
PFC, prefrontal cortex
POVM, positive operator value measurement
PSI, principle of selective influence
PSP, parameter space partitioning
QQ, quantum question
RKBS, reproducing kernel Banach space
RKHS, reproducing kernel Hilbert space
ROI, region of interest
RPE, reward prediction error
RSM, ratio scale matrix

- RT, response time
SNpc, substantia nigra pars compacta
SPM, statistical parametric map
SPRT, sequential probability-ratio test
SRW, simple random walk
STDP, spike-timing-dependent plasticity
SVM, support vector machine
TMS, transcranial magnetic stimulation
VTA, ventral tegmental area

Preface

Volume 2 of the New Handbook of Mathematical Psychology (NHMP) continues our goal to emphasize mathematical foundations and mathematical themes in the psychological sciences rather than to emphasize empirical facts and specific competing models (see the preface to Volume 1). This second volume, subtitled Modeling and Measurement, focuses on areas of mathematics that are in major use in constructing formal models of cognitive phenomena as well as formal approaches to understanding the role of measurement in psychology.

The first five chapters in Volume 2 deal with probabilistic models for cognitive phenomena. In particular, the first four chapters show how the standard Kolmogorov measure-theoretic axioms of probability, random variables, and stochastic processes are employed to formalize cognitive models in a number of areas of psychology. In these chapters the standard axioms are presented along with derived concepts such as Markov processes (Chapters 1 and 2), martingales (Chapter 1), stochastic filtrations (Chapter 3), and the conditions needed to identify and test models (Chapter 4). Chapter 5 discusses a newer and increasingly popular approach to cognitive modeling by constructing probabilistic models using quantum probability axioms rather than the standard Kolmogorov axioms.

The final four chapters involve the foundational use of mathematics in areas of mathematical psychology that share connections to areas outside of psychology. Chapter 6 describes approaches that incorporate concepts from computational cognitive neuroscience; Chapter 7 takes up problems in voting theory, an area that overlaps with economics and political science; and Chapter 8 takes up problems in classification that arise in the area of machine learning in computer science. Finally, Chapter 9 presents the latest work on the concept of measure-theoretic meaningfulness, an area in foundations of measurement that overlaps with philosophy and the physical sciences. These four chapters mostly describe and use areas of mathematics that do not have a probabilistic character. In particular, concepts in geometry are used in Chapters 7 and 9, function spaces (Hilbert and Banach) are used in Chapter 8 (also see Chapter 5), and abstract algebra is seen in several of these chapters, especially Chapter 9.

Chapter 1 by Diederich and Mallahi-Karai discusses probabilistic models of decision-making. Such models are developed using the mathematics of stochastic processes, and the chapter provides a very complete coverage of a variety of stochastic processes that have been and are being used to model decision-making. Many of the models discussed involve discrete-time and continuous-time Markov

processes; for example, discrete random walk models and continuous time diffusion processes such as the Wiener process and the Ornstein–Uhlenbeck process. The chapter shows how these diffusion processes can arise from limiting forms of the discrete models. Both uni- and multidimensional versions of these processes are formalized along with computational tools for dealing with them. In addition, other properties of stochastic processes such as stopping times and martingales are presented.

Chapter 2 by Jones takes up probabilistic models for binary decision-making, which are also discussed extensively in Chapter 1. Central in both of these chapters is the focus on probabilistic models that can account both for choice probabilities and the time it takes to make a response. Jones’ chapter focuses on models that assume that the decision to choose one of the items in a presented pair is based on temporal accumulation of evidence samples. While the chapter discusses both discrete and continuous random walk models, a major effort is made to represent these models in terms of Bayesian inference theory. Particularly important is an extension of earlier work (e.g., Jones & Dzhafarov, 2014) that analyzes the effect of incorporating intertrial variability in the model parameters. This assumption is the main way for diffusion models to predict observable differences in processing times for correct and incorrect decisions. However, the models that employ inter-trial variability have failed to provide a compelling rationale for their assumptions, and without a principled specification such models can predict any pattern of data. In other words, in the sense of falsifiability in Chapter 4, they cannot be rejected on data.

Chapter 3 by Houpt, Townsend, and Jefferson concerns properties of models for accounting for the time to complete a task consisting of multiple subcomponents, each of which must be processed to completion for the task itself to complete. Viewed generally, there are two general modeling schemes for such problems – serial processing and parallel processing of the components – and often models of both kinds can account for the same data (e.g., Townsend, 1972). The chapter provides a formal specification of serial and parallel models for such tasks cast in traditional Kolmogorov probability theory and the associated area of stochastic processes. These formalisms are presented along with the concept of filtration in stochastic processes. A key to the work discussed in the chapter is the way that selective influence is incorporated to compare models, namely where an environmental variable can selectively affect specific subprocesses in the task.

Chapter 4 by Doignon, Heller, and Stefanutti takes up important statistical issues concerning the scientific evaluation of parametric cognitive models for behavioral data. Such models have statistical parameters, and one important issue is whether a model is identified in the sense that different settings of the parameters necessarily generate different probability distributions of the observable data. Identified models allow unique measurement of their cognitive parameters based on observed data, and such models can be used to pinpoint the processes that are behind performance differences between participant groups. A related statistical issue is the question of whether a model is falsifiable (testable), that is, if it is possible, in

principle, to observe data that a model cannot account for. Important work by Bamber and van Santen (2000) provided the formal definitions along with some mathematical tools to investigate these issues, and Chapter 4 greatly expands on these mathematical tools including the presentation of polytopes in multidimensional Euclidean space. The new tools are illustrated by their application to models from knowledge space theory (e.g., see Chapter 5 in the first volume; Doignon and Falmagne, 1999; or Falmagne and Doignon, 2011).

Chapter 5 by Busemeyer and Kvam provides mathematical tools and examples related to the probabilistic modeling of cognitive phenomena using formalisms from quantum theory. The past decade or so has seen an increasing use of principles from quantum theory in the specification of cognitive models. Busemeyer and co-workers have been major developers of this approach (e.g., Busemeyer, Wang, & Townsend, 2006). In order to explicate the formalisms of quantum probability, the authors present the necessary axioms of Hilbert spaces, leading up to linear operators, basis vectors, and tensor product spaces. Given this background, the quantum probability axioms are presented and contrasted with the classical probability theory. Then several examples are presented where cognitive models based on quantum formalisms provide successful explanations for cognitive phenomena that seem difficult to explain by more traditional cognitive models.

Chapter 6 by Ashby is about computational cognitive neuroscience (CCN). This area did not exist in the early days of mathematical psychology; however, since the 1990s recent advances in cognitive neuroscience have made this area attractive to modelers. Ashby was one of the first mathematical psychologists to embrace CCN models (e.g., Ashby *et al.*, 1998), and later he contributed to a mathematical understanding of fMRI data (e.g., Ashby, 2011). The chapter clearly shows how supplementing the usual behavioral data, e.g., response choices, response times, and confidence ratings, with detailed neurobiological data can greatly restrict a modeler's choices and consequently increase the scientific validity of a model. The chapter focuses on some important and well-developed models in neuroscience concerning such problems as single spiking neurons, models for firing rates, and models for learning such as Hebbian and reinforcement learning algorithms.

Chapter 7 by Saari concerns a problem in the area of decision-making, namely finding formal rules for aggregating the choices of a group of agents (voters, attributes, data sources) to generate a choice that best and most "fairly" represents the group. The problems raised in the chapter go back to the seminal work of Kenneth Arrow (e.g., 1951), where it was shown that seemingly reasonable axioms for aggregating voters' rank orders of options could not in general be satisfied. Following Arrow's work, a number of paradoxical examples have been given that appear to challenge the possibility of formalizing any satisfactory rule for fairly aggregating agents' choices. Saari has become a leading theorist in this area by laying bare the exact mathematical reasons behind the paradoxes in voting theory, (e.g., Saari, 1995) and this chapter focuses on new aggregation rules that involve derived or stated paired comparison choices. Unlike other chapters that develop probabilistic models of decision-making, Saari's chapter mostly uses algebraic and geometric methods to develop its results.

Chapter 8 by Zhang and Zhang discusses the problem of classification (categorization) at an advanced mathematical level involving Hilbert spaces and Banach spaces. The problem of classification is to induce a classifier from a finite set of classified instances (exemplars) that will perform well for other new instances. This is clearly an ill-posed problem because there are always an unlimited number of schemes that can perfectly classify a given finite set of exemplars. Many cognitive models have been developed to delineate classifiers that can predict data in categorization experiments. In the area of statistical machine learning, a number of approaches have been developed which are quite separate from the cognitive modeling work. A major approach in machine learning assumes that possible classifiers are elements in a Hilbert function space. The chapter develops related systems of classifiers by dropping the Hilbert space requirement of an inner product, and this leads the authors to develop related properties in a suitable Banach space. The relevant mathematics of these spaces along with the needed formalism to address the classification problem are presented in detail. Of special interest is that the standard cognitive models are also discussed, and their connection to specific versions of the machine learning models is explicated.

Chapter 9 by Falmagne, Narens, and Doble provides the latest foundational work on the concept of meaningfulness. Meaningfulness is an important subtopic of foundations of measurement (abstract measurement theory), and this area has been a major topic in mathematical psychology since its beginnings in the 1950s, (e.g. Krantz *et al.*, 1971; Narens, 1985; and Suppes & Zinnes, 1963). Meaningfulness concerns the logical status of propositions and lawful relationships involving measured quantities under permissible scale transformations of the quantities. Meaningfulness was first introduced in the context of psychophysics in a seminal article by Stevens (1946), and formal interest in meaningfulness was greatly stimulated by the article by Luce (1959), where he proposed that the measurement scales of variables in a psychophysical law placed formal restrictions on what functional forms (logarithmic, power function, etc.) are possible. Since Luce's article, the meaningfulness problem has been analyzed primarily by utilizing the concept of homomorphisms in abstract algebra and relating meaningfulness to dimensional analysis in physics and Klein's (1872) Erlanger program in geometry. The measure-theoretic concept of meaningfulness has been criticized by some authors (e.g., Guttman, 1971; Michell, 1986), especially in relation to dimensional analysis and the nature of dimensional constants in physics (Dzhafarov, 1995). The primary value of Chapter 9 for this volume, however, is in the mathematical themes it cogently introduces, and this value does not depend on possible views of the scientific status of its central concepts.

References

- Arrow, K. J. (1951). *Social choice and individual values*. New York, NY: Wiley.
Ashby, F. G. (2011). *Statistical analysis of fMRI data*. Cambridge, MA: MIT Press.

- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105, 442–481.
- Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, 44, 20–40.
- Busemeyer, J. R., Wang, Z., & Townsend, J. (2006). Quantum dynamics of human decision making. *Journal of Mathematical Psychology*, 50, 220–241.
- Doignon, J.-P., & Falmagne, J. Cl. (1999). *Knowledge spaces*. Berlin: Springer-Verlag.
- Doignon, J.-P., & Falmagne, J.-Cl. (2016). Knowledge spaces and learning spaces. In W. H. Batchelder, H. Colonius, E. N. Dzhafarov, & J. Myung (eds.), *New handbook of mathematical psychology* (pp. 274–321). Cambridge: Cambridge University Press.
- Dzhafarov, E. N. (1995). Empirical meaningfulness, measurement-dependent constants, and dimensional analysis. In R. D. Luce, M. D'Zmura, D. Hoffman, G. J. Iverson, & A. K. Romney (eds.), *Geometric representations of perceptual phenomena* (pp. 113–134). Mahwah, NJ: Erlbaum.
- Falmagne, J.-C., & Doignon, J.-P. (2011). *Learning spaces*. Berlin: Springer-Verlag.
- Guttman, L. (1971). Measurement as structural theory. *Psychometrika*, 36, 329–347.
- Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121, 1–32.
- Klein, F. (1872). *Vergleichende Betrachtungen über neuere geometrische Forschungen. Verlag on Andreas Deichert, Erlangen*. (Available online at the University of Michigan Historical Mathematics Collection.)
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Volume 1: Additive and polynomial representations*. New York, NY: Academic Press.
- Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, 66, 81–95.
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, 100, 398–407.
- Narens, L. (1985). *Abstract measurement theory*. Cambridge, MA: MIT Press.
- Saari, D. G. (1995). *Basic geometry of voting systems*. New York, NY: Springer-Verlag.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, New Series, 103 (No. 2684), 677–680.
- Suppes, P., & Zinnes, J. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (eds.), *Handbook of mathematical psychology, Vol. 1*. New York, NY: John Wiley & Sons.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25, 168–199.

1 Stochastic Methods for Modeling Decision-making

Adele Diederich and Keivan Mallahi-Karai

1.1	Introduction	2
1.2	Probabilistic Modeling in Decision Theory	3
1.2.1	Information Accrual	4
1.2.2	Random Walk Models – An Example	5
1.3	Markov Chains	8
1.4	Markov Property	9
1.5	Examples of Markov Chains with a Finite State Space	11
1.6	Transition Probabilities in Several Steps	14
1.7	Distribution of Markov Chains: Computational Aspects	15
1.8	Markovianization: Markov Models with Bounded Memory	17
1.9	Stopping Times	18
1.10	Absorbing Markov Chains	20
1.11	Hitting Probabilities	23
1.11.1	More General Random Walk on Integers	24
1.12	Simple Random Walk with Absorbing Boundaries	26
1.13	Random Walks on the Line and the Grid	28
1.13.1	Simple Random Walk on \mathbb{Z}^2	28
1.13.2	Simple Random Walk on $\mathbb{Z}^d, d \geq 3$	30
1.14	Random Walk on Trees and its Applications	30
1.14.1	Applications of the Random Walk on the Tree	32
1.15	Continuous-time Markov Chains	33
1.16	Random Walks and the Wiener Process	35
1.17	The Distribution of the Wiener Process	38
1.17.1	The Wiener Process with Drift: Two Approaches	39
1.17.2	Functional Relation between the Parameters	41
1.18	Computation with the Wiener Process	42
1.18.1	Simulation of the Wiener Process	45
1.19	Multidimensional Wiener Process	46
1.20	Gaussian Vectors	46
1.20.1	The Covariance Matrix	47
1.20.2	Invariance of Gaussian Vectors under Affine Transformations	47
1.21	The Bessel Process	48
1.22	The Ornstein–Uhlenbeck Process	49
1.22.1	The Ehrenfest Urn Model	50

1.22.2 Construction of the Ornstein–Uhlenbeck Process	51
1.22.3 The Ornstein–Uhlenbeck Process: From Discrete to Continuous	53
1.23 Martingales and Their Applications	54
1.23.1 Martingales and Stopping Times	56
1.24 Decision-making with Multiple Attributes	58
1.24.1 Time and Order Schedules	59
1.24.2 Implementation of Random Schedules	65
References	69

1.1 Introduction

Decision-making situations, small or large, are everywhere: from simple and fast perceptual decisions to those with long-term effects, such as accepting or declining a job offer – we have to make decisions on a daily basis. The choice alternatives may be few or many; the outcomes may be certain or probable; the decision-maker may or may not be under time constraints. They all call for an attempt to find ways to model the underlying choice processes that can account for the data that can be collected during these processes. The goal is to come up with models that are, on the one hand, accurate and with explanatory power, and on the other hand, enjoy simplicity and robustness.

Mathematics has a role to play here. During the last 60 years mathematical models of decision theory have been developed to study these processes using tools from the theory of stochastic processes. These models have become the dominant approach to modeling decision processes in psychology and cognitive science. The probabilistic framework is theoretically consistent with the idea that the outcome of the choice and the time it takes to make a decision, i.e., the response time or reaction time, are not set or determined at the outset, but are something that emerges out of the deliberation process.

Intuitively a stochastic process is an entity that evolves randomly in time or space. In physics or biology and neuroscience this entity may be particles or neural activation. In psychology and cognitive science, depending on content and context, it is referred to as amount of information, activation, pieces of evidence, preference, and the like.

In psychology, two major classes of stochastic processes have mainly been applied to account for choice frequencies and choice response times. One class of models assumes that evidence for one option is at the same time evidence against the alternative option. They are mostly applied to binary choice situations. Within this class, random walk models accumulate discrete evidence in discrete time, whereas diffusion models accumulate continuous evidence in continuous time. The most commonly used version of the diffusion model is the Wiener diffusion model that linearly accumulates evidence without any decay (Ratcliff, 1978), but other models include the Ornstein–Uhlenbeck model that linearly accumulates evidence with decay (Busemeyer & Townsend, 1993; Diederich, 1995), and the leaky competing accumulator (LCA) model (Usher & McClelland, 2001) that

nonlinearly accumulates evidence with decay. The other class of models consists of accumulator and counter models. They can easily be extended to more than two choice alternatives, because an accumulator/counter is established for each choice alternative separately, and evidence is accumulated in parallel. A decision is made as soon as one counter wins the race to reach one preset criterion. The accumulators/counters may or may not be independent. Poisson-counter models are prominent examples, but random walk and diffusion models, one process for each alternative with a single criterion (absorbing boundary) for each process, can also be employed. Other accumulator models such as LATER (linear approach to threshold with ergodic rate) (Carpenter & Williams, 1995) and LBA (linear ballistic accumulator) (Brown & Heathcote, 2005) assume a deterministic linear increase in evidence for one trial. Randomness in responses occurs by assuming a normal distribution across the linear accumulation rate. These models are not considered further here.

In the following we focus on random walk/diffusion models. Our focus will not be on presenting the state-of-the-art research from the mathematical perspective or the latest debates in psychology and neuroscience. Rather, we will try to provide the underlying mathematical ideas and tools that have proven to be very successful in the last 60 years in psychological research. Along the way, we will discuss various models that have been proposed and try to lay out the underlying mathematical assumptions that have been placed. In particular, we start with a very simple model, a random walk, and develop more elaborate models from there. The focus will be on Markov chains, which has been referred to as the matrix approach to derive predictions of the models. Numerous examples provide deeper insight to the interplay between psychology and mathematics, i.e., mathematical psychology.

1.2 Probabilistic Modeling in Decision Theory

Sequential sampling models are among some of the most developed decision-making models. These models assume that characteristics of the choice options can be mapped onto a hypothetical numerical dimension representing the instantaneous level of information, activation, evidence, or preference. Further, they assume some random fluctuation of this value over time in the course of the accumulation process. Therefore, sequential sampling models can be built using stochastic processes, that is, a collection of random variables, representing the evolution of some system of random values over time,

$$\{X_\alpha\}_{\alpha \in A}.$$

Note that the *index set* or *parameter set* of the random variable X may be discrete ($\alpha = n \in \mathbb{N}$) or continuous ($\alpha = t \in \mathbb{R}^+$) and relates to the time of a realization of the random variable. We will call this set here *time set* or *time space*. The *state space* S defines the values or *states* that the random variables X_α can take on and may be also discrete (\mathbb{Z}) or continuous (\mathbb{R}). Let us make a brief remark on the connection between discrete and continuous models. Discrete-time models have

the advantage that their analysis often does not involve hard-core mathematical technology. This simplicity adds to their appeal, especially to the less mathematically sophisticated user. On the other hand, some experience with both of these set-ups is enough to show that the continuous processes are often more amenable to analysis. When at all possible, it is more helpful to derive closed formulas for continuous processes, and even in the absence of such formulas, one can often use various techniques from numerical analysis to return to the discrete set-up and use it as an approximation of the continuous process.

Note that there is often a close connection between discrete and continuous processes. The celebrated Wiener process can be viewed as the limit of suitably scaled random walks. This connection can also sometimes be used to analyze one of these models in terms of the other.

1.2.1 Information Accrual

During the last decades, various probabilistic models have been developed to explain the process of decision-making based on accumulation of evidence for the alternatives at hand. Each model represents the space of available information up to a certain point in time as a process that takes values in a subset of the Euclidean space \mathbf{R}^d , where the dimension d is often equal to the number of alternatives.

The process starts at a given point that represents the initial information or bias the decision-maker may have toward one of the choice alternatives. The information accrued up to time T (which may be discrete or continuous) corresponds to a point in this space. Thus, the entire deliberation process can be viewed as driven by a discrete or continuous random process. The discrete models often lead to models based on random walks, whereas continuous models are based on a number of diffusion models. The structure and parameters of this stochastic process highlight the underlying assumptions about the process one wants to investigate, for instance, sensory receptions, storing, memory retrieval, preference, categorizing, and more.

Most of the models that have been developed in these fields are based on Markov processes. The characteristic feature of a Markov process is that the dynamics of the process is determined by the current state of the process.

One of the implications of the Markov property is that the past can only influence the future through the present state of the system. Another way of stating this is that given the present state of the process, its future is independent from its past. This clearly indicates that a Markov process has a weak memory. This drawback aside, the Markov assumption allows one to bring a large body of existing mathematical theory (including tools from linear algebra and analysis) to bear on various applications. There are various generalizations of the Markov property that allow for some amount of memory. These systems that have a bounded amount of memory can also be recast as a Markov model with a different state space (see subsection 1.8). On the other hand, it must be mentioned that non-Markovian models (such as self-avoiding random walks) have proven to be much more difficult to analyze.

Another component of any decision model is the termination criterion. The stochastic process is run until a stopping criterion is satisfied. There are various ways to stop the process. The most obvious one is to have a fixed period for deliberation, after which the process is stopped. This may be related to an experimental set-up, in which the decision-maker is asked to make a response at a predetermined point in time. In other words, it is a time-based criterion. In many models, however, the criterion is satisfied when the process crosses a certain threshold, representing a sufficient amount of evidence in favor of one of the alternatives. Here the decision-maker sets an evidence-based criterion. The former criterion has also been called *fixed stopping rule*, the latter *optional stopping rule* (Busemeyer & Diederich, 2002). In the presence of more than two alternatives, it seems that both the accumulation process and the stopping criterion can be rather involved. In modeling situations in which decision-making is subject to time pressure, the criterion for stopping can also change with time. These will be discussed in the section on stopping times. We will start with a brief and rather informal discussion of the random walk model.

1.2.2 Random Walk Models – An Example

In this subsection, we will give an informal definition of random walks on the line and explain how this model can be used as a prototype of a decision-making system. More details will be given in the subsequent sections. We will start with an example of a *simple random walk with two absorbing boundaries*. A random walk $\{X_n\}$ is a stochastic process with discrete time space, that is, the realization of the random variable occurs at discrete times

$$t = 0, 1, 2, \dots$$

and discrete state space S . The random process X_n can thus take value in the set $\{0, \pm 1, \pm 2, \dots\}$. The model as such has few applications in psychology. However, it serves as an intuitive example and introduces basic concepts used throughout the chapter.

Suppose a person has to make a decision between two choice options A and B . At any moment in time, the person may sample information ξ for choosing A or B . For the sake of concreteness, let us assume that the probability of sampling information in favor of option A equals 0.3; the probability of sampling information ξ in favor of option B is 0.7. Amounts of information are coded in units of one. We will represent the total amount of gathered information by a real number. Assume that information in favor of option A amounts to moving from point u to the point $u + 1$ and information in favor of B amounts to moving from u to the point $u - 1$. Suppose that ξ denotes the unit of gathered information. We will express this by writing

$$\Pr[\xi = +1] = 0.3, \quad \Pr[\xi = -1] = 0.7.$$

The coin-tossing random variables ξ_i are examples of the *signed Bernoulli distribution*. A random variable ξ is said to have a signed Bernoulli distribution if there

is $0 < p < 1$ such that ξ takes values 1 or -1 with probabilities p and $1 - p$. We express this mathematically by writing

$$\Pr[\xi = 1] = p, \quad \Pr[\xi = -1] = 1 - p.$$

Signed Bernoulli random variables are variations of the ordinary Bernoulli random variables, where, instead of 1 and -1 , the random variable takes values of 1 and 0.

Two important assumptions have been made here: identical distribution of steps, and their independence.

Assumption 1.1 The random variables ξ_i are *identically distributed*. In other words, the probability that the information accrued at time i is ± 1 is the same as the probability that the information at time j be equal to ± 1 :

$$\Pr[\xi_i = 1] = \Pr[\xi_j = 1] = \frac{1}{2}, \quad \Pr[\xi_i = -1] = \Pr[\xi_j = -1] = \frac{1}{2}.$$

This assumption essentially states that the information source has reached a stationary distribution. Such assumptions are rather common in modeling problems.

Assumption 1.2 The random variables ξ_i are *independent*. In other words, the information accrued at time i will not influence any ξ_j for $j \neq i$. This can be mathematically expressed as follows. For any values of $\epsilon_1, \dots, \epsilon_n = \pm 1$, we have

$$\Pr[\xi_1 = \epsilon_1, \dots, \xi_n = \epsilon_n] = \prod_{1 \leq j \leq n} \Pr[\xi_j = \epsilon_j].$$

Let X_n denote the amount of information accumulated up to time unit n , $n \in \mathbb{N}$, i.e.

$$X_n = \sum_{i=1}^n \xi_i. \tag{1.1}$$

Continuing with our example, assume that at the beginning of the trial no amount of information has been sampled yet, $X_0 = 0$, and that the sampling process stops as soon as a critical value is reached for either initiating a response in favor of option A or in favor of option B . Let us assume that the critical value for option A is $\theta_A = 4$ and for option B is $\theta_B = -4$. The state space for this example is therefore given by

$$S = \{0, \pm 1, \pm 2, \pm 3, \pm 4\}.$$

The critical values $\{-4, 4\}$, a subset of the state space, are called *absorbing states*; the remaining intermediate states $S^* = \{-0, \pm 1, \pm 2, \pm 3\} \subset S$ are called *transient states*. Eventually the accumulated information leaves the transient states and is captured by one of the absorbing states. That is, the probability of absorption is 1.

The states and the probabilities for moving up or down, that is, the *transition probabilities* (defined properly later), can be conveniently presented in matrix form (Equation (1.2)). The rows of the matrix display all states, in our example $-4, -3, \dots, +3, +4$, and are represented by rows 1, 2, \dots , 9, respectively; and

similarly for the columns. This matrix (the entries within the brackets) is called the *transition probability matrix* and is denoted by \mathbf{P} .

$$\mathbf{P} = \begin{array}{c|cccccccccc} index & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline state & -4 & -3 & -2 & -1 & 0 & +1 & +2 & +3 & +4 \\ \hline 1 & -4 & \left[\begin{array}{ccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .7 & 0 & .3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .7 & 0 & .3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 \end{array} \right] \\ 2 & -3 & \\ 3 & -2 & \\ 4 & -1 & \\ 5 & 0 & \\ 6 & +1 & \\ 7 & +2 & \\ 8 & +3 & \\ 9 & +4 & \end{array} \quad (1.2)$$

Each cell, p_{ij} , of this matrix represents the transition probability of going from state s_i to state s_j . For example, row 5 reflects the transition probabilities from the neutral state (state 0) to either the state one step up (to state +1 from column 5 to column 6 with probability .3); or remaining in neutral (with probability 0); or to the state one step down (to state -1 from column 5 to column 4 with probability .7). The remaining rows are defined in a similar manner.

Three possible *realizations* of the information accumulation process X , also called *sample paths* or *trajectories*, are

- trial 1 (0, +1, 0, -1, -2, -3, -4, -3, -4)
- trial 2 (0, +1, +2, +3, +2, +3, +4)
- trial 3 (0, -1, 0, -1, -2, -1, -2, -3, -4, -3, -4)

In the first trial, the first amount of information gives evidence for choosing option A and the state changes from neutral to a value that moderately favors A (+1). However, the evidence is not sufficiently strong to decide. The next amount of information favors B , producing an evidence step back to neutral. The process continues until one of the critical values is reached, at which point the respective response is chosen, i.e., as soon as $X_n = \theta_A = +4$ or $X_n = \theta_B = -4$. In the case of trial 1, the evidence is sufficiently strong to make a decision in favor of B after the eighth sample. In the second trial there was enough information accumulated in favor of B after the sixth sample; in the last trial there was sufficiently strong evidence to make a decision in favor of B after the tenth sample. For the last example, the cumulative amount of information at each time is $X_0 = 0; X_1 = -1; X_2 = 0; X_3 = -1; X_4 = -2; X_5 = -1; X_6 = -2; X_7 = -3; X_8 = -4; X_9 = -3; X_{10} = -4$.

In the subsequent sections we will provide many relevant mathematical concepts that are related to the probabilistic modeling of decision-making. In any modeling problem, it is of paramount importance to underline the mathematical underpinnings of the model. This is necessary to see exactly how our assumptions about the decision-making processes are interpreted in mathematical formalism.

This will also help to find the pitfalls of such models. Apart from a rigorous mathematical theory, we will also discuss tools that can be helpful for practitioners. Most of these tools are standard to mathematicians, but may not be as well known to math users, hence their inclusion in this chapter. In the next section we will discuss the notion of Markov chain, which is the set-up for most discrete decision-making models.

1.3 Markov Chains

In this section we will formalize and extend the ideas discussed in subsection 1.2.2. As indicated above, in many modeling problems, one deals with sequences of random quantities

$$X_1, X_2, \dots, X_n, X_{n+1}, \dots$$

whose values are revealed with the time progress. The most basic situation is about a sequence of *independent* quantities. This means that a full knowledge of X_1, \dots, X_n does not influence the distribution of the variable X_n , or more precisely,

$$\Pr [X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_1 = s_1] = \Pr [X_n = s_n]. \quad (1.3)$$

Here we use the notation $\Pr [A|B]$ to denote the conditional probability¹. Note that the equality (1.3) indicates that the extra information provided by $X_{n-1} = s_{n-1}, \dots, X_1 = s_1$ has no bearing on predicting the value of X_n . The proverbial tossing of coins is perhaps the most well-known example of such processes. The sequence ξ_1, ξ_2, \dots discussed in the previous section is an example of such a sequence.

The theoretical simplicity is one of the appeals of such processes. However, in many real-world situations, one needs to deal with processes in which the distribution of the quantity X_n at time n somehow depends on that of the history of the process. As a motivating example, consider the following situation. Let P_d denote the price of a unit of the stock of company C on the d th day of the year. It is intuitively clear that the sequence P_1, P_2, \dots does not behave as a sequence of independent quantities. In fact, one expects the value P_{n+1} to be highly correlated with the values of P_m , at least for values of m close to n , even though this dependence is not deterministic. In other words, the knowledge of the value of the stock on days $d = 1, \dots, n$ is relevant for predicting its value on day $d = n + 1$. Here is one way of formulating this property. We assume that the value X_n of the sequence at time n is given by

$$X_{n+1} = F(X_n, \dots, X_{n-k}, S_n),$$

where S_n is a *random seed*, that is, a source of randomness independent of the sequence X_m . For instance, in the example of the random walk described in the previous section, we have

¹ Recall that the conditional probability of A given B is defined by the ratio $\Pr [A|B] = \Pr [A \cap B] / \Pr [B]$ for $\Pr [B] > 0$.

$$X_{n+1} = X_n + \xi_{n+1}.$$

In this example, the step taken at time $n + 1$ is the random seed involved in determining the position of the chain at time $n + 1$.

Note that we have postulated that the values of X_1, \dots, X_{n-k-1} do not directly affect the value of X_{n+1} . In other words, the system has a *limited memory*.

A particularly important case is when $k = 1$. In this case, X_n depends on X_{n-1} and the random seed S_n , but does not depend on the “older history” consisting of the values X_{n-1}, \dots, X_1 . A succinct (although not precise) description is that the only part of history relevant to predicting future is the current state. In the next section we will make this notion more precise.

1.4 Markov Property

We will begin with the formal definition of Markov chains. We will then proceed to some examples.

Definition 1.1 Let S be a finite set. A sequence X_1, X_2, \dots of random variables which take values in S is called a Markov chain if for all $n \geq 1$ and all $s_1, \dots, s_n \in S$ we have

$$\mathbf{Pr}[X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_1 = s_1] = \mathbf{Pr}[X_n = s_n | X_{n-1} = s_{n-1}]. \quad (1.4)$$

This condition is often referred to as the *Markov property*. Let us compare (1.4) to (1.3). Suppose that the elements of S correspond to the states of a system that evolves with time, in that we think of X_n as the *state of the system at time n*. Note that we have implicitly assumed that the time is discrete. Later we will consider analogous situations in which the time is assumed to be continuous. In (1.3), the entire history of the past is considered to be irrelevant for the prediction of the future. In (1.4), however, we are assuming that the value of X_n may depend on the past through its current value. Note that the information given by

$$X_{n-1} = s_{n-1}, \dots, X_1 = s_1$$

describes the full history of the system up to time $n - 1$. The Markov property simply states that the distribution of X_n only depends on its past inasmuch as it determines the value of X_{n-1} . In other words, the history of the system prior to time $n - 1$ is only relevant through the value of X_{n-1} . As the system evolves with time, all the history except for the current state is erased. To study a Markov chain we will need to work with the transition probabilities defined next:

Definition 1.2 The *transition probabilities* of the Markov chain (X_n) are defined by

$$p_{ij}^{[n]} = \Pr[X_n = j | X_{n-1} = i].$$

In most of the applications in decision theory, we work with Markov chains that are *time-homogeneous*. This condition means that the transition probability $p_{ij}^{[n]}$ does not depend on the time n in which the transition is happening. In such cases, the superscript n is suppressed and the shorthand notation p_{ij} will be used.

Given a time-homogeneous Markov chain, a systematic way of keeping track of the transition probabilities is to put them in an $N \times N$ matrix, where N denotes the number of states. From now on, we will always describe a time-homogeneous Markov chain with N states s_1, \dots, s_N by an $N \times N$ matrix

$$\mathbf{P} = (p_{ij})_{1 \leq i, j \leq N}$$

where p_{ij} denotes the transition probability of the chain from state s_i to state s_j . The matrix \mathbf{P} defined above is called the *transition matrix* of the Markov chain. For instance, p_{ii} is the probability that the Markov chain stays put, given that it is at state s_i . In the following theorem, we sum up two of the fundamental properties of the transition matrix.

Theorem 1.3 *The transition matrix \mathbf{P} of any Markov chain satisfies the following two properties:*

1. *The transition probabilities are non-negative, that is, for all $1 \leq i, j \leq N$,*

$$p_{ij} \geq 0.$$

2. *Each row adds up to 1, that is, for $1 \leq i \leq N$, we have*

$$\sum_{j=1}^N p_{ij} = 1.$$

The first property is obvious. Let us explain the second property. Note that the entries in the i th row of \mathbf{P} indicate the probabilities of making a transition from a fixed state s_i to one of the other states s_j , where j could vary. Because exactly one of these transitions will happen, the law of total probability implies that the sum of respected probabilities is 1, that is

$$\sum_{j=1}^N p_{ij} = \sum_{j=1}^N \Pr[X_n = j | X_{n-1} = i] = 1.$$

A matrix with these two properties is called a *stochastic matrix*. So, the transition matrix of any Markov chain is a stochastic matrix. Conversely, it is easy to see that if \mathbf{P} is a stochastic matrix, one can define a Markov chain with transition matrix \mathbf{P} . Note that the same cannot be said about the columns. In some of the examples below we will encounter Markov chains where the sum of entries of a column is not equal to 1.

1.5 Examples of Markov Chains with a Finite State Space

In this section we will give several examples of Markov chains.

Example 1.4 The simplest non-trivial Markov chain consists of two states that will be denoted by s_1 and s_2 . The transition matrix of this Markov chain is given by

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix},$$

where the parameters p and q are arbitrary probabilities satisfying $0 \leq p \leq 1$ and $0 \leq q \leq 1$. Note that the cases $p = 0$ or $q = 0$ behave somewhat differently from the case $0 < p, q$. In the latter case, the Markov chain will flip back and forth from one state to the other indefinitely. If, for instance, $p = 1$ or $q = 1$, then the Markov chain will be absorbed in one of the states. In mathematical models of decision-making, we often work with Markov chains that either naturally demonstrate an absorbing behavior, or the chain will be stopped, for instance, by using a stopping time.

Example 1.5 (Ehrenfest urns) Consider n balls placed into two urns. We assume that the initial number of balls in the urns is given by $[k : n - k]$, which indicates that there are k balls in the first and $n - k$ balls in the second urn. At each stage, one ball is selected randomly and is moved to the other urn. Clearly, the set of all configurations of the system (or the states of the Markov chain) is given by

$$S = \{s_i = [i : n - i] : 0 \leq i \leq n\}.$$

Note that there are $N = n + 1$ states in this chain. For $i = 0$, we clearly have $p_{01} = 1$ and for $i = n$, we have $p_{n,n-1} = 1$. For $1 \leq j \leq n - 1$, we have

$$p_{i,i-1} = \frac{i}{n}, \quad p_{i,i+1} = \frac{n-i}{n}.$$

The transition matrix of this Markov chain for $n = 4$ is given by

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Clearly, the number of balls in the urns continues to fluctuate with a tendency from the urn that has more balls to the urn with fewer balls. This model can be used for modeling a system where the distance of the system to its equilibrium determines the size of the restoring force.

Example 1.6 (Lazy Ehrenfest) Let us now consider a variation of the Ehrenfest chain. In the modified version, the chain stays put with probability $1/2$. When it moves, it moves according to the same rule as before, i.e., one of the balls is

randomly chosen and moved to the other urn. It is easy to see that the transition matrix for this chain is given by

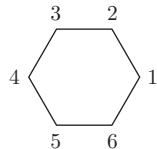
$$\mathbf{P}_{\text{lazy}} = \frac{\mathbf{I} + \mathbf{P}}{2} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/6 & 1/2 & 1/3 & 0 \\ 0 & 1/3 & 1/2 & 1/6 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

Remark (Lazy version of a Markov chain) If \mathbf{P} is the stochastic matrix associated to a Markov chain, then

$$\mathbf{P}_{\text{lazy}} = \frac{\mathbf{I} + \mathbf{P}}{2}$$

is also the stochastic matrix that corresponds to a *lazy* version of the original chain, where \mathbf{I} denotes the identity matrix. The lazy walk forces the chain to stay put with probability $1/2$. The rest of the probability is divided proportionately between the states of the chain according to the initial probabilities in matrix \mathbf{P} . In some theoretical contexts, it is more advantageous to work with the lazy version of a Markov chain. In a similar fashion, one can consider the transition matrices $\mathbf{P}_\epsilon = \epsilon\mathbf{I} + (1 - \epsilon)\mathbf{P}$, where the “laziness” constant $0 \leq \epsilon \leq 1$ can be chosen arbitrarily. Using lazy versions of Markov chains, one can slow down their movement.

Example 1.7 (Random walk on a cycle) Consider the Markov chain with n states, which are represented by the nodes on a cycle. The chain moves from one node to one of the neighboring nodes. Let us assume that the probability of moving clockwise is p and the probability of moving counterclockwise is $q = 1 - p$.



The transition probabilities are given by

$$p_{ij} = \begin{cases} p & \text{if } j = i + 1 \\ 1 - p & \text{if } j = i - 1 \\ 0 & \text{otherwise.} \end{cases}$$

The transition matrix of this chain for $n = 6$ is given by

$$\mathbf{P} = \begin{pmatrix} 0 & p & 0 & 0 & 0 & q \\ q & 0 & p & 0 & 0 & 0 \\ 0 & q & 0 & p & 0 & 0 \\ 0 & 0 & q & 0 & p & 0 \\ 0 & 0 & 0 & q & 0 & p \\ p & 0 & 0 & 0 & q & 0 \end{pmatrix}$$

An interesting feature of this transition matrix is that each column also adds up to 1. Stochastic matrices with this property are called *doubly stochastic*.

Example 1.8 (Random walk on a line segment with absorbing boundaries) Consider a Markov chain with the state space given by

$$S = \{s_0, s_1, \dots, s_L\},$$

where $L > 0$ is a fixed integer. The initial state of the chain is $X_0 = a$ with $0 \leq a \leq L$. At every move, the chain shifts one step to the right or to the left with probabilities p and $1 - p$, with the exception that when the chain reaches s_0 or s_L it will stop. For $L = 4$, the transition matrix of this chain is given by

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Note that s_0 and s_L are the absorbing boundaries of this chain.

Example 1.9 (Random walk on a line segment with bouncing boundaries) Let us modify the Markov chain in the previous examples as follows. As before, we assume that the chain moves one step to the right or left with probabilities p and $1 - p$. But now, we assume that the states s_0 or s_L are bouncing, that is, the chain moves from s_0 to s_1 and from s_L to s_{L-1} . The transition matrix \mathbf{P} for $L = 4$ must be modified as follows:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

In spite of the similarity between these chains, their long-term behavior is radically different.

Note that we can also implement a random walk with one absorbing and one bouncing boundary. This captures the process in which only one response is possible, for instance, in simple detection tasks.

Example 1.10 Consider a sequence of trials, where the outcome of each trial can be success or failure. Suppose that success happens with probability p and failure with probability $1 - p$. A success run of length r is a sequence of r successes happening in a row. Let X_n denote the success run at time n . If the $n + 1$ -st trial results in a success, then $X_{n+1} = X_n + 1$, and if it results in failure, then $X_{n+1} = 0$. Hence, the transition matrix of this Markov chain has the following form:

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 & 0 & \dots \\ 1-p & 0 & p & 0 & \dots \\ 1-p & 0 & 0 & p & \dots \\ 1-p & 0 & 0 & 0 & \dots \\ \vdots & & & & \end{pmatrix}$$

Example 1.11 A coupon collector wishes to collect a set of n different types of coupons. Each acquired coupon has probability $1/n$ of being one of these coupons. Let X_k denote the number of coupons collected up to time k . Clearly $X_0 = 0$. Because the first collected coupon is always new, we have $X_1 = 1$. The second coupon will be new with probability $1 - \frac{1}{n}$ and it will be the same coupon as the first one with probability $1/n$. More generally, it is easy to see that if i coupons have been collected up to time k , then the probability that the $k+1$ -st coupon is new is $\frac{n-i}{n}$. This implies that the transition probabilities for $i < n$ are given by

$$p_{i,i+1} = \frac{n-i}{n}, \quad p_{i,i} = \frac{i}{n}, \quad p_{i,j} = 0, \quad j \neq i, i+1.$$

Clearly, once all coupons have been collected the process stops moving:

$$p_{n,n} = 1.$$

This means that the state s_n is an absorbing state. The transition matrix for $n = 4$ is as follows:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

1.6 Transition Probabilities in Several Steps

Consider a Markov chain with transition probabilities p_{ij} . We are interested in the probability

$$p_{ij}^{(n)} = \Pr[X_{n+k} = j | X_k = i].$$

In other words, we want to calculate the probability that the chain makes a transition from state i to state j in n steps. We also set $p_{ij}^{(0)} = 1$ if and only if $i = j$ and 0 otherwise. For transition in one step, we clearly have

$$p_{ij}^{(1)} = p_{ij}.$$

The following theorem shows how these probabilities can be calculated inductively.

Theorem 1.12 (Chapman–Kolmogorov) *Suppose i and j are states of a Markov chain with N states. Then*

$$p_{ij}^{(n+1)} = \sum_{1 \leq k \leq N} p_{ik} p_{kj}^{(n)}.$$

Let us first explain the reason behind the statement. Any transition from s_i to s_j in $n + 1$ steps would require the chain to move to a state s_k in the first move. This event has probability p_{ik} . Once carried out, the probability of moving from k to j in the remaining n steps is given by $p_{kj}^{(n)}$. Now, using the Markov property, the second event is independent of the first move, given the state of the chain after the first move. From here the statement follows. Let us now give a precise proof. We will argue by induction on n . It is clear that the statement holds for $n = 0$. Suppose that the statement holds for $n - 1$. Then we have

$$\begin{aligned} \mathbf{Pr}[X_{n+1} = j | X_0 = i] &= \sum_{k \in S} \mathbf{Pr}[X_{n+1} = j, X_1 = k | X_0 = i] \\ &= \mathbf{Pr}[X_1 = k | X_0 = i] \mathbf{Pr}[X_{n+1} = j | X_0 = i, X_1 = k]. \end{aligned} \tag{1.5}$$

In the last equality, we used the general fact

$$\mathbf{Pr}[A \cap B | C] = \mathbf{Pr}[B | C] \mathbf{Pr}[A | B \cap C],$$

which readily follows from definition. Using the Markov property, one can easily see that

$$\mathbf{Pr}[X_{n+1} = j | X_0 = i, X_1 = k] = \mathbf{Pr}[X_{n+1} = j | X_1 = k],$$

from which the statement follows.

Theorem 1.12 can be most effectively used when it is formulated using matrix formalism.

Corollary *Let \mathbf{P} denote the transition matrix of a Markov chain. Then for any $1 \leq i, j \leq m$, and any $n \geq 1$, the transition probability from the state s_i to state s_j in n steps is given by*

$$p_{ij}^{(n)} = (\mathbf{P}^n)_{ij},$$

where \mathbf{P}^n is the n th power of the transition matrix \mathbf{P} . In other words, the probability $p_{ij}^{(n)}$ is given by the ij entry of the matrix power \mathbf{P}^n .

1.7 Distribution of Markov Chains: Computational Aspects

Consider a Markov chain X_0, X_1, \dots with the state space $S = \{s_1, \dots, s_m\}$. One of the implications of the Markov property is that if we know the distribution of X_0 and the transition matrix \mathbf{P} , we can compute the distribution of X_n .

Suppose π is a $1 \times m$ row vector with non-negative entries that add up to one:

$$\pi = (\pi_1, \dots, \pi_m).$$

One can view π as a probability vector representing the distribution of X_0 , that is, the initial distribution of the chain. This has a clear psychological interpretation. A decision-maker starts with an initial bias. In the most extreme case, it could be that the initial state is deterministically chosen, that is, $\pi_i = 1$ for some i and $\pi_j = 0$ for $j \neq i$. More generally, π is any distribution on the set of states of the chain. Depending on whether the decision-maker's initial preference state is biased towards an alternative, the values of π_i can be set. The following theorem will be used quite often in the future.

Theorem 1.13 *Suppose that the initial distribution of a Markov chain with the transition matrix \mathbf{P} is given by the row vector π . Then the distribution of the position of the chain at time n , that is, X_n , is given by the vector $\pi\mathbf{P}^n$.*

Let us illustrate the power of this by giving an example. Recall the Markov chain with two states discussed in Example 1.4. We denote the states by s_1 and s_2 . The transition matrix is thus given by the matrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

where $0 < p < 1$ and $0 < q < 1$. Suppose that the initial distribution is given by $\pi = (\pi_1, \pi_2)$. The distribution of the states after n steps will be given by $\pi\mathbf{P}^n$. So the problem boils down to computing the n th power \mathbf{P}^n . This can certainly be done by a computer algebra system. Here, we will show an alternative approach for doing this. This approach is based on the method of diagonalization, that is, finding an invertible matrix U and a diagonal matrix D such that

$$\mathbf{P} = UDU^{-1}. \tag{1.6}$$

Let us assume that 1.6 holds. It can be easily checked that then $\mathbf{P}^n = U D^n U^{-1}$. As D is a diagonal matrix, D^n can be computed efficiently, which in turn allows us to compute \mathbf{P}^n . The entries on the diagonal matrix D are the eigenvalues of \mathbf{P} . In this case, they are given by

$$\lambda_1 = 1, \quad \lambda_2 = 1 - p - q.$$

The corresponding eigenvectors can also be computed:

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} p \\ -q \end{pmatrix}.$$

It is now easy to check that if

$$U = \begin{pmatrix} 1 & p \\ 1 & q \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix},$$

then

$$\mathbf{P} = UDU^{-1}.$$

From here we have

$$\begin{aligned}\mathbf{P}^n &= UD^nU^{-1} = \begin{pmatrix} 1 & p \\ 1 & q \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda_2^n \end{pmatrix} \begin{pmatrix} 1 & p \\ 1 & q \end{pmatrix}^{-1} \\ &= \frac{1}{q-p} \begin{pmatrix} q - p\lambda_2^n & p(\lambda_2^n - 1) \\ q(1 - \lambda_2^n) & -p + q\lambda_2^n \end{pmatrix}.\end{aligned}$$

1.8 Markovianization: Markov Models with Bounded Memory

As we have seen so far, a Markov chain is a stochastic process in which the future is independent of the past, when conditioned on the present. There are, however, situations in which the dynamics of the process may also depend on the past in a way that is not reflected in the present. In other words, having access to the recent history of the process may be useful in providing a more accurate description of the process. Such models have been used in risk management, cloud data mining, and one can imagine them to be useful in decision theory as well. Let us give a formal definition of this notion.

Definition 1.14 Let S be a finite set. A (discrete-time) Markov chain with memory m with the state space S is a stochastic process $X_t, t = 1, 2, \dots$ such that for all $n \geq m$ and all $s_1, \dots, s_n \in S$ we have

$$\begin{aligned}\Pr[X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_1 = s_1] \\ = \Pr[X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_{n-m} = s_{n-m}].\end{aligned}\quad (1.7)$$

Fixing m , let us call the values of X_{n-1}, \dots, X_{n-m} the recent history of the process. The idea of a Markov chain with bounded memory is that the memory has a buffer of size m , so that the recent history can be directly used (and not just through the present state of the system as in the case of Markov chains) in determining the distribution of the future. However, the old past, consisting of X_t for $t < n - m$, can only indirectly (through its influence on the recent memory) have an impact.

Markov models with bounded memory offer an additional flexibility in modeling various processes. On the other hand, their analysis can be reduced to the analysis of an extended Markov chain. We will describe this process, which is also known as Markovianization, in detail. For each $t \geq m$, we will define

$$Y_t = (X_{t-1}, X_{t-2}, \dots, X_{t-m}).$$

Note that Y_t captures the recent history of the process. Let us now define a new state space S^* by the m -fold product of the original state space S . In other words, an element of S^* is a vector of length m of the form

$$\mathbf{s} = (s_1, \dots, s_m)$$

where $s_i \in S$ for $1 \leq i \leq m$. One can then easily see that the process Y_t with the state space S^* satisfies the Markov property. Let us illustrate these ideas by an example.

Example 1.15 Consider a random walk on the line with the constraint that the random walker cannot make three moves to the left one after the other. Suppose X_n denotes the state of the chain at time n , and ξ_1, ξ_2, \dots denotes the steps taken by the random walker. The assumption of the problem is that

$$\Pr[\xi_n = -1, \xi_{n-1} = -1, \xi_{n-2} = -1] = 0.$$

Note that (X_n) is not a Markov chain. To see this, assume that $X_8 = 6$. This can be realized by either one of the following scenarios:

$$\begin{aligned} \text{Scenario 1 : } & \xi_1 = \xi_2 = \cdots = \xi_6 = 1, \quad \xi_7 = \xi_8 = -1 \\ \text{Scenario 2 : } & \xi_1 = \xi_2 = -1, \quad \xi_3 = \cdots = \xi_8 = 1. \end{aligned} \tag{1.8}$$

Note that in Scenario 1, the last two steps have been to the left, so the walker in the next step can only go to 7. On the other hand, in Scenario 2, the last step has been to the right, so the random walker is free to go to 5 or 7. This means that the knowledge of the current state ($X_8 = 6$) is not sufficient to determine the probability of the next step. However, as indicated above, we can make a new Markov chain out of this process. In order to do this, we keep track of the last three steps of the chain. In other words, we expand the state space to consist of all (i, j, k) of integer vectors with $|j - k| = |i - j| = 1$. This is the *expanded process* denoted by Y_n . Here, $Y_n = (i, j, k)$ means that the random walker was at point i at time $n - 2$ and is at step j at time $n - 1$. We can now easily verify that Y_n is indeed a Markov chain whose transition matrix can be simply computed.

1.9 Stopping Times

Another important concept from probability theory which is used in decision theory is *stopping time*. Suppose that $\{X_t\}_{t \geq 0}$ is a stochastic process. Here, the index variable t can be discrete, that is, run over non-negative integers, or it can be continuous, in which case, it runs over non-negative real numbers. A stopping time is a random variable τ with the property that for any $t \geq 0$, the event $\tau \leq t$ can be determined on the basis of values X_s for $s \leq t$. In other words, a stopping time is a rule for stopping which can be determined only on the basis of the current (and possibly previous) values of the process.

To make this idea clear, we will start with a simple example, followed by a non-example. Suppose X_t , for $t = 1, 2, \dots$, denotes the outcomes of throwing a fair die. Let τ_1 denote the first time that a 6 appears. Mathematically, we write

$$\tau_1 = \min\{t : X_t = 6\}.$$

Note that the event $\tau_1 \leq k$ is equivalently described by $X_j = 6$ for some $1 \leq j \leq k$. This implies that knowing whether the chain has stopped before time k , equivalently, $\tau_1 \leq k$, can be determined solely based on the knowledge of X_1, \dots, X_k . Hence τ_1 is a stopping time.

Let us now consider a variation of τ_1 . Let τ_2 denote the step one step before the first 6 arrives. In other words:

$$\tau_2 = \min\{t : X_t = 6\} - 1.$$

In this case, $\tau_2 \leq k$ iff $X_j \leq k$ for some $j \leq k + 1$. In other words, determining whether $\tau_2 \leq k$ will require some “peeping into the future”, which is not allowed in the definition of stopping times. This implies that τ_2 is not a stopping time.

Stopping time τ_1 is an example of a family of stopping times called *hitting time*. For instance, the stopping criterion can be set to be the first entry or the first exit time. Suppose \mathcal{A} is a subset of the set of all states. Define the first entry and the first exit time associated with \mathcal{A} as:

$$\tau_{\mathcal{A}}^{\text{ent}} = \min\{t : X_t \in \mathcal{A}\}.$$

$$\tau_{\mathcal{A}}^{\text{ext}} = \min\{t : X_t \notin \mathcal{A}\}.$$

The stopping times $\tau_{\mathcal{A}}^{\text{ent}}$ and $\tau_{\mathcal{A}}^{\text{ext}}$ are called, respectively, the *first passage time*, and the *first exit time*. Some of the stopping times of the utmost importance in decision theory are of these kinds. As an example, suppose that X_t is a Wiener process (for the definition, see Section 1.16) and $z > 0$ is a given threshold. Set $\tau = \min\{t \geq 0 : X_t = z\}$. When X_t stands for the accumulated amount of information up to time t , the stopping time τ indicates when X_t passes a threshold z .

The example above can be modified as follows. Suppose

$$\tau = \min\{t \geq 0 : X_t = \pm z\}.$$

This stopping time can be used in the following situation. In a binary decision process with alternatives A and B , suppose X_t denotes the accumulated information for A . When X_t hits z , enough information has been accumulated in favor of A . On the other hand, when $X_t = -z$, the information against A , and hence in favor of B , has been accumulated.

The random variable X_{τ} is usually referred to as the *stopped process*. One of the first important questions that naturally arises is understanding the distribution of the stopped process. For instance, by definition, for any subset \mathcal{A} of the set of states, the stopped process $X_{\tau_{\mathcal{A}}^{\text{ent}}}$ takes values in \mathcal{A} . Hence, its distribution is an \mathcal{A} -valued random variable. In applications to decision theory, where \mathcal{A} is often identified with the set of alternatives, the probabilities

$$\Pr[X_{\tau_{\mathcal{A}}^{\text{ent}}} = a]$$

are often called *choice probabilities*.

Various statistical properties of τ such as $\mathbf{E}[\tau]$, $\text{Var}[\tau]$ correspond to quantities of interest in applications. Similarly, the distribution of the stopped variable X_{τ} is

useful. For instance, in the aforementioned example, X_τ is clearly a Bernoulli random variable which can take values of $\pm z$. However, determining the probabilities $\Pr[X_\tau = \pm z]$ is more challenging. The next example shows that in some special cases, one can compute statistical properties of the stopping times.

Example 1.16 Consider the coupon collector problem from Example 1.11. Assuming that $X_0 = 0$, denote by τ the time at which the state s_n is reached. Then

$$\mathbf{E}[\tau] = n \sum_{k=1}^n \frac{1}{k} \approx n \log n.$$

This example illustrates a useful technique that one can use in computing the expected value of stopping times. The idea is to split τ into a sum of “simpler” stopping times. In this example, we denote by τ_k the time it takes for the first k distinct coupons to be accumulated. It is obvious that

$$\tau = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1}).$$

Let us now consider $\tau_k - \tau_{k-1}$. It takes τ_{k-1} steps for the first $k-1$ coupons to be collected. From this point on, every coupon is new with probability $p = \frac{n-k+1}{n}$. Because each collected coupon is independent of the previous ones, this means that $\tau_k - \tau_{k-1}$ has a geometric distribution with parameter p . This implies that

$$\mathbf{E}[\tau_k - \tau_{k-1}] = \frac{1}{p} = \frac{n}{n-k+1}.$$

Adding these up, we have

$$\mathbf{E}[\tau] = \sum_{k=1}^n \mathbf{E}[\tau_k - \tau_{k-1}] = n \sum_{k=1}^n \frac{1}{n-k+1} = n \sum_{k=1}^n \frac{1}{k}.$$

It is well known that the harmonic series can be approximated by the associated integral:

$$\sum_{k=1}^n \frac{1}{k} \approx \int_1^n \frac{1}{t} dt = \log n,$$

from which the claim follows.

1.10 Absorbing Markov Chains

In most of the applications in decision theory, we will deal with absorbing Markov chains. For simplicity, let us assume that the set of the states of the Markov chain is finite: $S = \{s_1, \dots, s_n\}$.

Definition 1.17 We call a state s_i of a Markov chain *absorbing* if

$$p_{ii} = 1.$$

Note that this implies $p_{ij} = 0$ for all $j \neq i$.

We denote the set of absorbing states by S_a and the rest of the states by S_r . It is convenient to reorder the states of the Markov chain into absorbing and non-absorbing states. Assume that there are m non-absorbing and $n - m$ absorbing states. Note that if s_i and s_j are both absorbing states, then

$$p_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Ordering the states in such a way that absorbing states are listed first, the transition matrix takes the following simple form:

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{R} & \mathbf{Q} \end{pmatrix}, \quad (1.9)$$

where \mathbf{Q} is an $m \times m$ matrix, and it governs the transition between non-absorbing states. The matrix \mathbf{R} has size $m \times n - m$ and governs the absorption, that is, moving from a non-absorbing to an absorbing state. As indicated above, once we are in an absorbing state, we will remain there with probability 1. This is the reason for the appearance of the identity matrix in the upper-left corner of \mathbf{P} . We will henceforth refer to this form as the *canonical form* of the transition matrix.

Example 1.18 Consider a process that starts at the point $(0, 0)$. At every period, one of the coordinates will increase by one unit. Assume that the probability that the first (second, respectively) coordinate will increase by 1 is p ($1 - p$, respectively). The process stops when one of the coordinates is 2. The state space of the process can be described as

$$S = \{(0, 0), (1, 0), (0, 1), (1, 1), (2, 0), (2, 1), (1, 2), (0, 2)\}.$$

Note that the absorbing states are given by

$$S_a = \{(2, 0), (2, 1), (1, 2), (0, 2)\}.$$

When written in the canonical form, the matrices \mathbf{Q} and \mathbf{R} are given by

$$\mathbf{Q} = \begin{pmatrix} 0 & p & 1-p & 1 \\ 0 & 0 & 0 & 1-p \\ 0 & 0 & 0 & p \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ p & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-p \\ 0 & p & 1-p & 0 \end{pmatrix}.$$

Assume now that \mathbf{P} is the transition matrix of a Markov chain with some absorbing states in the canonical form. It is not difficult to see that powers of \mathbf{P} can be computed in the following straightforward way:

$$\mathbf{P}^k = \begin{pmatrix} \mathbf{I} & 0 \\ (I + \mathbf{Q} + \dots + \mathbf{Q}^{k-1})\mathbf{R} & \mathbf{Q}^k \end{pmatrix}.$$

As $\mathbf{Q}^k \rightarrow 0$ as $k \rightarrow \infty$, we see that the matrix

$$\mathbf{N} = I + \mathbf{Q} + \mathbf{Q}^2 + \dots$$

is well-defined. This matrix is called the *fundamental matrix* of the absorbing chain.

Theorem 1.19 *For a non-absorbing state s_i and an absorbing state s_j , let b_{ij} denote the probability that the chain absorbs in s_j if it starts at s_i . Then $\mathbf{B} = (b_{ij})$ is given by*

$$\mathbf{B} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R}.$$

Proof Assume the chain has started at s_i . With probability p_{ij} the chain will be absorbed in j . If this does not happen, the chain has moved to another non-absorbing state and from there eventually gets absorbed in s_j . This allows us to write:

$$b_{ij} = p_{ij} + \sum_{s_k \in S-A} p_{ik} b_{kj},$$

which is equivalent to the matrix equation

$$\mathbf{B} = \mathbf{R} + \mathbf{Q}\mathbf{B},$$

which implies the result. \square

Corollary *The probability that the chain finally absorbs is equal to 1.*

We can also state a more precise version of this formula. Suppose a is one of the absorbing states, for instance, associated with choosing alternative A , and $n \geq 1$. In order to start the chain, we will also need an initial distribution. This is given by the vector \mathbf{Z} , here viewed as a row vector, representing the probabilities for the starting position(s) of the process over the transient states before the decision process begins. The probability that the chain lands in a after n steps is given by

$$\Pr_{\mathbf{Z}} [\tau_a = n] = \mathbf{Z} \mathbf{Q}^{n-1} \mathbf{R}_a, \quad n = 1, 2, \dots, \infty, \quad (1.10)$$

where \mathbf{R}_a is the column of the matrix \mathbf{R} containing the transition probabilities leading to the absorbing state a .

From here, we can also compute the probability of absorption at a , given that the chain starts with the distribution \mathbf{Z} on the set of non-absorbing states. If we denote this event by $Abs(a)$, then

$$\Pr_{\mathbf{Z}} [Abs(a)] := p_A = \mathbf{Z} \sum_{n=1}^{\infty} \mathbf{Q}^{n-1} \mathbf{R}_a = \mathbf{Z} (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R}_a. \quad (1.11)$$

It is also useful to compute the moments of τ_a , given that the chain is absorbed at a . These are given by

$$\mathbf{E} [\tau_a^r | Abs(a)] = \frac{1}{p_A} \mathbf{Z} \sum_{n=1}^{\infty} n^r \mathbf{Q}^{n-1} \mathbf{R}_a. \quad (1.12)$$

In particular, the mean time equals

$$\mathbf{E} [\tau_a | \text{Abs}(a)] := ET_a = \frac{1}{p_a} \mathbf{Z}(\mathbf{I} - \mathbf{Q})^{-2} \mathbf{R}_a. \quad (1.13)$$

Alternatively, one can ask about the expected value and variance of the time to absorption, given the starting point of the chain. In this direction, we have

Theorem 1.20 *Let $\hat{\tau}_i$ be the random variable that gives the number of the periods it takes for the process starting from s_i to absorb (so, by definition, $\tau_i \geq 1$.) Then*

$$\mathbf{E} [\hat{\tau}_i] = \frac{1}{1 - p_{ii}}, \quad \text{Var} [\hat{\tau}_i] = \frac{p_{ii}}{(1 - p_{ii})^2}.$$

Notice that $\mathbf{E} [\hat{\tau}_i]$ tends to be large when p_{ii} is close to 1, that is, when the chain is likely to stay put at state s_i . Similarly, when $p_{ii} \approx 0$, the variance $\text{Var} [\hat{\tau}_i]$ is close to zero. This implies that if the probability of leaving s_i is close to 1, then the amount of time spent before absorption takes place is almost deterministic.

1.11 Hitting Probabilities

Absorption probabilities studied in the previous subsection are special cases of hitting probabilities. Suppose s_i and s_j are two states of a Markov chain. Define the probability $f_{ij}^{(n)}$ as the probability that starting from s_i , the first visit to j will occur exactly after n periods. In other words,

$$f_{ij}^{(n)} = \Pr [X_{n+1} = j, X_m \neq j \text{ for } 1 \leq m \leq n | X_1 = i].$$

We will also set

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}.$$

In order to avoid confusion, let us point out that $p_{ij}^{(n)}$ is the probability that the chain will be at s_j after n steps given that it started from the state s_i . Clearly, f_{ij} denotes the probability of reaching s_j if the chain starts from s_i . This does not preclude the possibility that s_j has been visited prior to time n . In particular, $p_{ij}^{(n)} \geq f_{ij}^{(n)}$.

In order to compute these probabilities, we will relate them to the transition probabilities. The probability $p_{ij}^{(n)}$ can be computed by conditioning on the first passage time:

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}.$$

In particular, we have

Theorem 1.21 *For any state i of a Markov chain we have*

$$p_{ii}^{(n)} = \sum_{j+k=n} p_{ii}^j f_{ii}^k.$$

An efficient way of encoding all of these equalities is to use generating functions. Consider the generating functions defined by

$$F(z) = \sum_{n=1}^{\infty} f_{ii}^{(n)} z^n$$

$$U(z) = \sum_{n=1}^{\infty} p_{ii}^{(n)} z^n$$

This just means that we consider two power series where the coefficients of z^n are precisely $f_{ii}^{(n)}$ and $p_{ii}^{(n)}$, and try to find a relation between them. In fact, using the above equation it is easy to see that

$$U(z) = 1 + U(z)F(z).$$

From here it follows that

$$F(z) = \frac{U(z) - 1}{U(z)}.$$

There are many situations in which $U(z)$ admits a closed form. In such cases, one can use the above formula to compute the probabilities $f_{ii}^{(n)}$. The following example involves random walk on integers which will be discussed later.

Example 1.22 For the random walk on the set of integers $S = \{\dots, -1, 0, 1, \dots\}$, we have

$$U(z) = \sum_{n \geq 0} \frac{(2n)!}{n!^2} 2^{-2n} z^n = (1 - z^2)^{-1/2}.$$

This implies that

$$F(z) = 1 - (1 - z^2)^{1/2} = \sum_{n \geq 1} \frac{(2n)!}{(2n-1)(n!)^2} 2^{-2n} z^{2n}.$$

Extracting the coefficient of z^{2n} gives

$$f_{00}^{(2n)} = \frac{(2n)!}{(2n-1)(n!)^2} 2^{-2n}.$$

We will now discuss in detail random walks on the set of integers \mathbb{Z} . Often, we will assume that the random walk starts at the origin, that is, $X_0 = 0$. At each point, the random walker moves to the right with probability p and to the left with probability $q := 1 - p$. Let us review some of the basic properties of this random walk.

1.11.1 More General Random Walk on Integers

Recall that a Markov chain is a sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ which satisfies the *Markov property*. This means that the probability of moving to state s_k depends only on the present state s_j and not on the previous states, i.e.,

$$p_{ij} := \mathbf{Pr}[X_{n+1} = s_j | X_0, X_1, \dots, X_n = s_i] = \mathbf{Pr}[X_{n+1} = s_j | X_n = s_i].$$

In this subsection, we are interested in more general random walks on the set of integers. The state space of this chain is indexed by the set of integers:

$$S = \{\dots, s_{-2}, s_{-1}, s_0, s_1, s_2, \dots\},$$

and the transition probabilities are given by

$$p_{ij} = \mathbf{Pr}[X_{n+1} = s_j | X_n = s_i] = \begin{cases} p & \text{if } j = i + 1 \\ q & \text{if } j = i - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (1.14)$$

Let us examine some of the assumptions underlying this model for information accumulation. First, note that this process is *one-dimensional*. In other words, all the information accumulated by the decision-maker is encapsulated into one number. Another feature of this process is geometric: the transition only happens from one state to a neighboring state, that is, there are no big jumps. This is particularly desirable, because it underlines the fact that the information accumulation process is continuous and gradual.²

Note that the position of the chain at time n is the sum of the moves in the first n steps. In other words, we can write

$$X_n = \xi_1 + \dots + \xi_n$$

where each summand ξ_i is equal to ± 1 . Recasting X_n in this form has many advantages. First, note that ξ_i , $1 \leq i \leq n$ are independent and identically distributed random variables, and

$$\mathbf{Pr}[\xi_i = 1] = p, \quad \mathbf{Pr}[\xi_i = -1] = q.$$

Later in this section we will use this property to compute the mean and variance of the position of the X_n . Before proceeding, note that a special case of the random walk, that is when $p = q = \frac{1}{2}$, is of special importance. This special case corresponds to an unbiased walk, and has some properties that are different from other walks. Let us now turn to the computations.

$$\mathbf{E}[X_n] = \mathbf{E}\left[\sum_{i=1}^n \xi_i\right] = n\mathbf{E}[\xi_1] = n(1 \cdot p + (-1) \cdot q) = n(p - q). \quad (1.15)$$

Also, because $\text{Var}[\xi_1] = \mathbf{E}[\xi_1^2] - \mathbf{E}[\xi_1]^2 = 4pq$, it follows from independence that

$$\text{Var}[X_n] = n\text{Var}[\xi_1] = 4npq. \quad (1.16)$$

Finally, the probability distribution of X_n , that is, the value k of the process at time $n \geq 1$, is

$$\mathbf{Pr}[X_n = k] = \binom{n}{(k+n)/2} p^{(k+n)/2} q^{(n-k)/2},$$

² Note, however, that recent work in neurophysiology doubts that. It is assumed that information accumulation occurs in jumps. We are not discussing this perspective here.

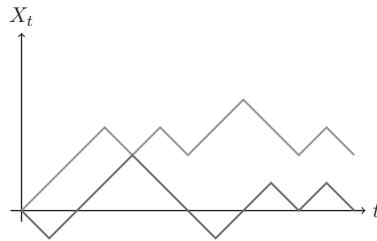


Figure 1.1 Two sample paths of the simple random walk.

with $n \geq |k|$, k and n being either both even or both odd. In fact, as X_n and n have the same parity, it follows that if n and k are of different parity, then $\mathbf{Pr}[X_n = k] = 0$. Otherwise, for $X_n = \xi_1 + \dots + \xi_n = k$, we require $\frac{k+n}{2}$ of the terms ξ_j for $1 \leq j \leq n$ to be equal to $+1$, and the rest to be -1 . The probability of this event is clearly given by the above formula. This shows that the probability that the accumulated evidence is at k after n steps is a binomial random variable.

In Figure 1.1, the graphs of two different realizations of the random walk on the set of integers have been plotted. The horizontal axis indicates time t , and the vertical axis is the location of the random walker.

At this point we will remark on a point which can be the source of much confusion. An important fact is that both paths can be realized for both the simple random walk with parameter $p = 1/2$ and a biased random walk with parameter $p = 0.2$. On the other hand, the probability that the upper path is realized in the first 12 steps of the simple random walk equals

$$p = (0.5)^6 \cdot (0.5)^6 = (0.5)^{12} \approx 0.00024.$$

The probability that the same path is realized by the biased walk with parameter 0.2 equals

$$p = (0.2)^6 \cdot (0.8)^6 \approx 0.004.$$

As one can see, the same path has a much larger probability of being realized when the parameter is 1/2.

1.12 Simple Random Walk with Absorbing Boundaries

Let us now consider the following modification of the walk, which is of interest in modeling. We will restrict the state space to

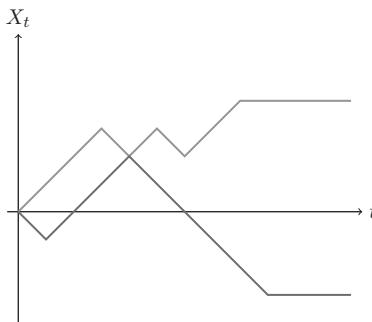
$$S = \{s_i : i \in \mathcal{I}\}, \quad \mathcal{I} = \{-m_B, -m_B + 1, \dots, 0, \dots, m_A - 1, m_A\} \subset \mathbb{Z}. \quad (1.17)$$

Note that S is finite and has $m = m_A + m_B + 1$ elements. The index i for $1 \leq i \leq m$ corresponds to the state $i - 1 - m_B$. The transition probabilities are now modified as follows.

For the set of transient, i.e., non-absorbing states S^* , the corresponding set is

$$\mathcal{I}^* = \{-m_B + 1, \dots, m_A - 1\}.$$

Set $\theta_A = m_A$ and $\theta_B = -m_B$.



Two quantities are of foremost interest to psychologists:

1. the probability that the process eventually reaches one of the decision boundaries for the first time (the criterion to initiate a response), called *first passage probability* (the probability of absorption);
2. the time it takes for the process to reach one of the boundaries for the first time, called *first passage time* (the time (steps) until absorption).

The former quantity is related to the observed relative frequencies, the latter usually to the observed mean choice response times or the observed choice response time distribution.

With $q = (1-p)$ and z as starting point ($X_0 = z$) the first passage probability (or absorption probability) for the simple random walk, i.e., the probability to choose option A , is given by (for derivations see Tuckwell, 1995, p. 135)

$$\Pr[A] = \frac{(q/p)^z - (q/p)^{\theta_B}}{(q/p)^{\theta_A} - (q/p)^{\theta_B}}, \quad p \neq q$$

and

$$\Pr[A] = \frac{z - \theta_B}{\theta_A - \theta_B}, \quad p = q.$$

The probability to choose option B is

$$\Pr[B] = \frac{(q/p)^{\theta_A} - (q/p)^z}{(q/p)^{\theta_A} - (q/p)^{\theta_B}}, \quad p \neq q$$

and

$$\Pr[B] = \frac{\theta_A - z}{\theta_A - \theta_B}, \quad p = q.$$

The mean number of steps taken until the walk is absorbed (i.e., the mean reaction time, RT) is the same for both choice option, i.e.,

$$\text{Mean } RT_{A/B} = \frac{\mathbf{Pr}[A](\theta_A - z) + \mathbf{Pr}[B](\theta_B - z)}{p - q}, \quad p \neq q$$

and

$$\text{Mean } RT_{A/B} = (z - \theta_B)(\theta_A - z), \quad p = q.$$

As an example, consider the case given by $z = 0$, $p = \mathbf{Pr}[\xi = +1] = 0.3$, $\theta_A = 4$ and $\theta_B = -4$. Then the probability to choose option A is $\mathbf{Pr}[A] = 0.0326$, to choose option B is $\mathbf{Pr}[B] = 0.9674$ and the mean RT for choosing any of the two options is 9.3473 time units.

1.13 Random Walks on the Line and the Grid

In this section, we will give a detailed construction of the random walk on the d -dimensional grid. Recall from the previous section that when $d = 1$, the random walk is described by a pair of real numbers p and q which are assumed to be positive and $p + q = 1$. Then, the Markov chain X_n with the state space $S = \mathbb{Z}$, starting at $X_0 = 0$ with the transition probabilities, is given by

$$p_{x,x+1} = p, \quad p_{x,x-1} = 1 - p, \quad x \in \mathbb{Z}.$$

There is a straightforward generalization of this to the d -dimensional grid. Although from a purely mathematical point of view, this is a natural generalization, it can also be applied to modeling problems in decision theory. Before we discuss these possible applications, let us describe this generalization in detail.

In this model, the state space $S = \mathbb{Z}^d$ consists of all points with integer coordinates in the d -dimensional space. The neighbors of a point $x \in S$ consist of the points y such that the vector $y - x$ has exactly one non-zero entry, which is equal to ± 1 . Note that as there are d directions, and in each direction one can go forward or backward, each point has exactly $2d$ neighbors. Let us introduce some notation to describe this generalization.

For $1 \leq i \leq d$, let e_i denote the vector with 1 in the i th entry and zero elsewhere. Also, for each $1 \leq i \leq d$ and each sign \pm , assume that $p(\pm e_i)$ are positive numbers which satisfy $\sum_{i=1}^d p(\pm e_i) = 1$. We will now consider a random walk on \mathbb{Z}^d with the transition probabilities given by

$$p_{x,x+e_i} = p(e_i), \quad p_{x,x-e_i} = p(-e_i), \quad x \in \mathbb{Z}^d.$$

The random walker following this law will move forward in direction i with probability $p(e_i)$ and move backward in direction i with probability $p(-e_i)$. The special case where $p(e_i) = p(-e_i) = \frac{1}{2d}$ for all $1 \leq i \leq d$ is called the *simple random walk* (SRW) on the d -dimensional grid.

1.13.1 Simple Random Walk on \mathbb{Z}^2

The simple random walk on \mathbb{Z}^2 is defined with the transition probabilities given by

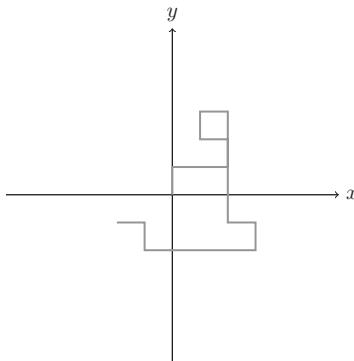


Figure 1.2 A sample path for the simple random walk in two dimensions.

$$\begin{aligned} p((x,y), (x+1,y)) &= p((x,y), (x-1,y)) = p((x,y), (x,y+1)) \\ &= p((x,y), (x,y-1)) = \frac{1}{4}. \end{aligned}$$

Figure 1.2 shows a realization of the simple random walk on \mathbb{Z}^2 .

Remark Note that, in general, the coordinates of a random walk in dimension d are dependent random variables. It is somewhat surprising that the simple random walk on \mathbb{Z}^2 is *isomorphic* to the product of two *independent* random walks on \mathbb{Z} . This can be seen as follows.

Consider a random walk defined by the process $W_n = (X_n, Y_n)$. As mentioned above, X_n and Y_n are not independent random walks. In fact, from $X_{n+1} - X_n = 0$ one infers that $Y_{n+1} - Y_n = \pm 1$. Let us introduce the new coordinates

$$U_n = X_n + Y_n, \quad V_n = X_n - Y_n.$$

By considering four possible cases, one can see that $U_{n+1} = U_n \pm 1$, and $V_{n+1} = V_n \pm 1$. Aside from this, each one of the four possibilities

$$U_{n+1} = U_n \pm 1, \quad V_{n+1} = V_n \pm 1$$

occurs with probability 1/4. For instance, $U_{n+1} = U_n + 1, V_{n+1} = V_n + 1$ happens precisely when $(X_{n+1}, Y_{n+1}) = (X_n + 1, Y_n)$, that is, the random walker has moved forward in the x -axis. From here, one can easily check that U_n and V_n are indeed independent random walks on \mathbb{Z} . The following example shows an application of this fact.

Example 1.23 We have already obtained the return probability of the simple random walk in dimension 1:

$$p_{00}^{(2k)} = \frac{1}{2^{2k}} \binom{2k}{k}.$$

As the random walker in dimension 2 returns to its starting point, precisely when this happens in both x and y coordinates, the above remark shows that in dimension 2 we have

$$p_{00}^{(2k)} = \left(\frac{1}{2^{2k}} \binom{2k}{k} \right)^2.$$

This implies that the random walk is recurrent in dimension 2. This means that for any point $(a, b) \in \mathbb{Z}^2$, the random walk will eventually visit this point. This fact is useful in defining certain stopping times.

1.13.2 Simple Random Walk on \mathbb{Z}^d , $d \geq 3$

The combinatorial problem in dimension $d = 3$ is a bit more involved. It is not hard to see that

$$p_{00}^{(2k)} = \frac{1}{6^{2n}} \sum_{i+j \leq k} \left(\frac{n!}{i!j!(k-i-j)!} \right).$$

The sum above is rather complicated, but can still be estimated. On the one hand, it is easy to see that the function $(x, y, z) \mapsto x!y!z!$ assumes its minimum (under the constraint that $x + y + z = k$) for $x = y = z = k/3$. Using Stirling's formula, we will see that when n is divisible by 3, we have the upper bound

$$p_{00}^{(2k)} \leq C_3 n^{-3/2}$$

One can easily show that the same inequality (with a slightly different constant C_3) also holds for all values of n . Now, as the series

$$\sum_{n=1}^{\infty} n^{-3/2}$$

diverges, we obtain the transience of the random walk in dimension 3.

More generally, it can be shown that in dimension d , the return probability is given by

$$p_{00}^{(2k)} \sim C_d n^{-d/2}.$$

As in the case $d = 3$, the random walk will be transient for all values of $d \geq 3$.

1.14 Random Walk on Trees and its Applications

Random walks on graphs are other Markov chains that have found applications in decision theory. Recall that a graph G consists of a set of vertices, denoted by $V(G)$, and a set of edges, denoted by $E(G)$. For two vertices $x, y \in V(G)$, we write $x \sim y$ when x, y form an edge. When $x \sim y$, we also say that x and y are connected. The set of vertices connected to x , also called the neighbors of x , is denoted by $N(x)$. For $x \in V(G)$, the number of vertices y with $x \sim y$ is called the

degree of x and is denoted by $\deg x$. Clearly $\deg x$ is the number of elements in $N(x)$.

In applications, a specific vertex x_0 of the graph is designated as the initial state. The simple random walk on G is a Markov chain with the state space $V(G)$. One then sets $X_0 = x_0$ and the transition probabilities are given by

$$p_{xy} = \begin{cases} \frac{1}{\deg x} & \text{if } x \sim y \\ 0 & \text{otherwise.} \end{cases}$$

Random walks can be defined in a more general way. Suppose that for each pair (x, y) of vertices of G with $x \in y$, a non-zero probability c_{xy} is assigned such that

$$\sum_{y \in N(x)} c_{xy} = 1.$$

One can view c_{xy} as the probability of making a transition from x to y in one step. Note that in general c_{xy} and c_{yx} do not have to be equal. One can then define a Markov chain where the probability of transition from x to y is given by c_{xy} .

Let us see an application of this to random walk on a rooted tree. A rooted tree has a distinguished vertex, called the *root*. Every vertex of the tree has exactly one “inward” neighbor. For simplicity, we will always assume that each vertex has the same number d of outward neighbors, although this is not necessary. There is a convenient way to label the vertices of a rooted tree. Assume that each vertex has out-degree $d + 1$. We will use the symbols $0, 1, \dots, d$ to label these neighbors. Let us use \emptyset for the root. The vertices in the first layer will be labeled by $0, 1, \dots, d$. The neighbors of vertex i in the first layer are labeled by $i0, \dots, id$. The labeling can be extended in a similar fashion. So, the label attached to a vertex in layer i is a sequence

$$z_1 z_2 \dots z_i$$

where the value of z_1, \dots, z_i can be any number from 0 to d . Let us describe the random walk on a rooted tree. Assume that $p(0), \dots, p(d)$ are non-negative real numbers adding up to 1. The initial point of the random walk is often assumed to be the root, that is, $X_0 = \emptyset$. We also assume that the random walker moves from layer i to layer $i + 1$ in each move. If $X_n = x_1 x_2 \dots x_n$ denotes the position of the random walker at time n , the position of the walker at time $n + 1$ is $X_{n+1} = x_1 x_2 \dots x_n x_{n+1}$, where x_{n+1} is any of the numbers $0, 1, \dots, d$, chosen with probabilities $p(0), \dots, p(d)$. From here, one can easily see that

$$\Pr [X_n = x_1 x_2 \dots x_n] = p(x_1) \cdots p(x_n).$$

Figure 1.3 shows the first few layers of a binary tree.

Note that the random walks considered here do not backtrack. This is consistent with using this Markov chain for modeling decision-making. Recall that many models of decision-making are based on an information sampling process. In such models, a counter is attached to each of the alternatives. The counters get updated once per unit of time, until one of them reaches a given threshold.

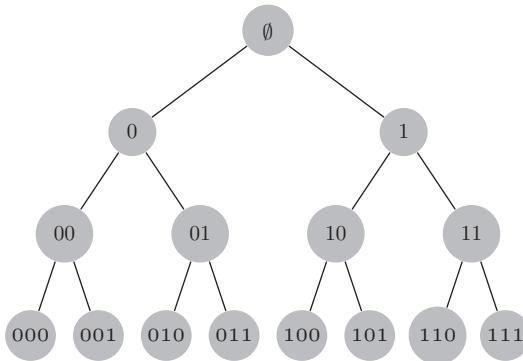


Figure 1.3 Three layers of a binary tree.

1.14.1 Applications of the Random Walk on the Tree

The random walk on the tree can be used to define a number of interesting processes. The simplest process that can be defined in this way is the simple random walk on \mathbb{Z} . Suppose X_1, X_2, \dots denote the outcome of a random walk on the tree. As an example, consider the segment

$$\emptyset, 1, 10, 101, 1011, 10110, 101100, 1011001, 10110011, \dots$$

For a sequence $x = x_1 \dots x_n$, let us define $\beta(x)$ as the number of occurrences of 1 minus the number of occurrences of 0. Now, consider $\beta(X_n)$. In the example at hand, the random walk on the tree will be transformed to

$$0, 1, 0, 1, 2, 1, 0, 1, 2, \dots$$

It is easy to see that the sequences obtained in this way describe a random walk on integers.

There are, however, more interesting processes that can be defined using the random walk on the tree. Fix $d \geq 1$, and let $x = x_1 \dots x_n$ be a sequence. Let us define by $\gamma(x)$ the number of occurrences of 1 minus the number of occurrences of 0 in the last d terms of the sequence. More precisely,

$$\gamma(x) = \#\{n - d + 1 \leq j \leq n : x_j = 1\} - \#\{n - d + 1 \leq j \leq n : x_j = 0\}.$$

For $d = 3$, we obtain the following sequence

$$0, 1, 0, 2, -2, -2, 2, \dots$$

The process defined here can be viewed as way of modeling a decision-making situation in which evidence is accrued with time, but due to the limitation in memory, only the last d units of information or evidence determine the state of the chain. Alternatively, this process can be described as follows: let ξ_1, ξ_2, \dots be signed Bernoulli random variables. For $j \leq 0$, set $\xi_j = 0$. For $n \geq 1$, define

$$X_n = \xi_n + \xi_{n-1} + \dots + \xi_{n-d+1}.$$

X_n will be the process defined in the previous paragraph. Note one important difference between this process and the usual random walk. X_n defined here is, by nature, a bounded random variable. In fact, $-d \leq X_n \leq d$ holds for all values of n . This is consistent with our assumption on the boundedness of memory.

1.15 Continuous-time Markov Chains

Although discrete-time random processes are somewhat more elementary objects, it turns out that it is more convenient to work with continuous analogues of them. A continuous-time process is a family $\{X_t\}$ of random variables which is parameterized by non-negative real numbers. In other words, for every $t \geq 0$, X_t is a random variable.

This definition is obviously too general. In most applications, we will require extra conditions on $\{X_t\}_{t \geq 0}$. For instance, it is often convenient to assume that X_t depends on t in a continuous way. Moreover, we need to impose restrictions on the distribution of X_t . The theory of continuous-time Markov processes has a lot of conceptual similarity to the discrete model, but due to the fact that time is now continuous, it needs to be set up differently. Recall that a Markov chain is a discrete stochastic process

$$X_0, X_1, X_2, \dots$$

where the distribution of X_{n+1} depends on the value of X_n . We are now going to deal with continuous time chains with a finite number of states. This implies that the transition from one state to another does not happen at integer times.

Definition 1.24 A continuous-time process X_t is called a Markov process if for any sequence $t_1 < t_2 < \dots < t_n$ of times and any states $s_1, \dots, s_n \in S$ the following equality holds

$$\Pr[X_{t_n} = s_n | X_{t_{n-1}} = s_{n-1}, \dots, X_{t_1} = s_1] = \Pr[X_{t_n} = s_n | X_{t_{n-1}} = s_{n-1}]. \quad (1.18)$$

There are two advantages in working with continuous time processes. First, note that real-world decision-making happens in continuous time, and hence continuous-time models offer a better chance of capturing essential features of the process. Aside from this, there are also some technical advantages in working with continuous models. For instance, periodicity, which sometimes occurs in discrete-time Markov chains (a random walker on the set of integers is always at an even point at an even time), completely disappears when we pass to continuous time.

On the other hand, as opposed to the discrete case where the probability distribution is captured by one matrix, in the continuous case one has to work with the generator of the Markov chain. Before we give the definition of the generator, let us recall the definition of transition probabilities.

Definition 1.25 The transition probabilities in a continuous-time Markov chain are defined by

$$p_{ij}(t_1, t_2) = \Pr[X_{t_2} = s_j | X_{t_1} = s_i].$$

The chain is called time-homogeneous or atemporal if $p_{ij}(t_1, t_2) = p_{ij}(0, t_2 - t_1)$, that is, when $p_{ij}(t_1, t_2)$ only depends on $t_2 - t_1$. In this case, we simply write $p_{ij}(t_2 - t_1)$ in lieu of $p_{ij}(t_1, t_2)$.

All the chains considered from now on are assumed to be time-homogeneous. Let us denote by \mathbf{P}_t the transition matrix of a continuous-time Markov chain. It is clear that \mathbf{P}_0 is the identity matrix, and the family \mathbf{P}_t satisfies the Chapman–Kolmogorov equation:

$$\mathbf{P}_{s+t} = \mathbf{P}_s \mathbf{P}_t, \quad s, t \geq 0.$$

As we are mostly interested in applications, we have dropped some of the technical conditions which are needed in the results below. In all likelihood, these conditions will be satisfied in most of the applications that arise in practice. Let us assume that the Markov chain has N steps.

Definition 1.26 There exists an $N \times N$ matrix \mathbf{G} with the following properties:

1. For all $1 \leq i \leq N$, we have

$$g_{ii} \leq 0.$$

2. For all **distinct** $1 \leq i \neq j \leq N$, we have

$$g_{ij} \geq 0.$$

3. For every $1 \leq i \leq N$, we have

$$\sum_{j=1}^N g_{ij} = 0.$$

Such that the transition probabilities satisfy the following differential equations:

1. (Forward equations) $p'_{ij}(t) = \sum_{k=1}^N p_{ik}(t)g_{kj}$.
2. (Backward equations) $p'_{ij}(t) = \sum_{k=1}^N g_{ik}p_{kj}(t)$. In other words, \mathbf{P}_t satisfies the differential equation

$$\mathbf{P}'(t) = \mathbf{G}\mathbf{P}_t = \mathbf{P}_t\mathbf{G}$$

and is hence given by

$$\mathbf{P}_t = e^{t\mathbf{G}}.$$

An intuitive way of thinking about \mathbf{G} is that g_{ij} gives the *rate* of passing from state s_i to state s_j .

Example 1.27 One of the most basic examples of continuous-time Markov processes is the birth process. Suppose $\lambda_1, \lambda_2, \dots$ are positive numbers and consider the generator matrix \mathbf{G} given by

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \dots \\ 0 & -\lambda_1 & \lambda_1 & \dots \\ 0 & 0 & -\lambda_2 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

This Markov chain takes values in the set of natural numbers, and is always non-decreasing. The parameters λ_n have the following probabilistic interpretation. The probability that the chain moves from state n to state $n + 1$ over the time interval $(t, t + h)$, where h is a small quantity, is roughly $\lambda_n h$. To be more precise, one has $p_{n,p+1}(h) = \lambda_n h + o(h)$, respectively, where $o(h)$ stands for a quantity which is of the order of magnitude less than h . This, in particular, implies that over a short duration h of time, with probability $\lambda_n h$, the process will stay at state n .

This process can be modified in the following way so that X_t can also decrease. To do this, we need to work with two sequences $\lambda_0, \lambda_1, \dots$ and μ_1, μ_2, \dots of non-negative numbers. We will also set $\mu_0 = 0$. The generator of the birth-death process is given by the following matrix

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Here, λ_n and μ_n are the tendencies for birth and death, respectively, when the population has size n . In other words, the probability that over a period h of time (as always h is a small quantity) the chain moves from n to $n + 1$ ($n - 1$, respectively) is given by $\lambda_n h + o(h)$ ($\mu_n h + o(h)$, respectively). The probability that the chain moves by more than one step is $o(h)$, and hence considerably smaller.

1.16 Random Walks and the Wiener Process

The most fundamental continuous-time random process is the Wiener process. This particular process not only occupies a privileged position in pure as well as applied mathematics, but also provides a basis for the construction of a large variety of other useful processes. The relevance of this process to the mathematical modeling of decision theory stems from the fact that the flow of information can be modeled by the movement of the Wiener process. As will be detailed below, one of the key defining characteristics of the Wiener process is that its increments over disjoint time intervals are independent random variables. When applied to this framework, it simply means that the amounts of information accumulated over disjoint time intervals are statistically independent.

Another ingredient needed to model decision-making is a criterion for stopping the process. Often, we assume that the decision-making takes place at the moment that the Wiener process hits some predefined boundary. In the psychological literature the boundaries are also called *decision criterion*, *decision boundary*, *decision threshold*, or simply *criterion*. As we will see, the Wiener process, the most applied sequential sampling process in psychology, can be considered as a scaled random walk.

A careful construction of this process requires some mathematical sophistication. Instead, we will opt for a less rigorous approach, which is self-contained, and instead focus on elucidating those aspects of the Wiener process that are related to applications we would like to present. We will start by defining the Wiener process. This definition spells out the key properties of the Wiener process.

Suppose that X_t denotes the net amount of information gathered up to time t in favor of an alternative in a decision-making situation. Suppose that the decision-maker starts the process at the initial point $X_t = x_0$. Let us denote by $p(x, t|x_0)$ the conditional probability density of the X_t . The fact that $p(x, t|x_0)$ is a density function implies the following two properties:

1. $p(x, t|x_0) \geq 0$.
2. $\int_{-\infty}^{+\infty} p(x, t|x_0) dx = 1$.

Note that

$$p(x, t + \Delta t) = \frac{1}{2}(p(x + \Delta x, t) + p(x - \Delta x, t)).$$

Definition 1.28 The standard Wiener process (also known as Brownian motion) is the process $\{W_t : t \geq 0\}$ defined to be the unique process which satisfies the following properties.

1. $W_0 = 0$.
2. $W_t - W_s$ is a Gaussian variable with distribution $N(0, t - s)$.
3. $\{W_t\}$ has independent increments, i.e., if $t_1 < \dots < t_r$ then the Gaussian random variables $W_{t_2} - W_{t_1}, \dots, W_{t_r} - W_{t_{r-1}}$ are independent.
4. W_t is a continuous function of t .

Remark The first hypothesis simply states that the Wiener process starts at zero. When dealing with processes with different initial values, one can consider process $X_t = W_t + c$.

Remark Note that for $t - s$ the distribution of the increment $W_t - W_s$ of the Wiener process is the normal distribution with mean zero and variance $t - s$. In particular, when t and s are close, $W_t - W_s$ has a small variance, and is thus almost deterministically zero.

Remark Note that the independent increment property of the Wiener process implies that for every $t_1 < \dots < t_n$, the vector $X_t = (W_{t_1}, \dots, W_{t_n})$ is a Gaussian vector. Note that for $s < t$ we can split W_t , because $W_t = W_s + (W_t - W_s)$ is the sum

of two independent Gaussian variables with variances s and $t - s$. This situation must be compared with a random walk, for which $X_n = X_m + (X_n - X_m)$ is the sum of two independent walks with variances m and $n - m$.

Theorem 1.29 *The Wiener process satisfies the following:*

1. For $s \geq 0$, we have $\mathbf{E}[W_s] = 0$.
2. For $s \geq 0$, we have $\mathbf{E}[W_s^2] = s$.
3. For $0 \leq s \leq t$, we have $\mathbf{E}[W_s W_t] = s$.

1 and 2 follow immediately from the definition. For 3, note that

$$\mathbf{E}[W_s W_t] = \mathbf{E}\left[W_s^2 + W_s(W_t - W_s)\right] = s.$$

There are various constructions for the Wiener process. In the most intuitive approach, it is defined as the limit of a scaled symmetric simple random walk. Let us explain this in more detail.

Consider once more the simple random walk with unit steps and state space S as in (1.17) with discrete time $t = 0, 1, 2, \dots$. Start with a time interval $[0, t]$ and divide it into n subintervals of length $\tau = t/n$. We will now modify the random walk so that the process makes a step at times $t = 0, \tau, 2\tau, 3\tau, \dots$. Further, instead of steps up and down of unit magnitude assume that the size of the steps is either $+\Delta$ or $-\Delta$. The state space is now

$$S = \{\dots, -2\Delta, -\Delta, 0, +\Delta, +2\Delta, \dots\}.$$

Furthermore, assume that the distribution of independent identically distributed random variables ξ_i is now given by

$$\mathbf{Pr}[\xi_i = +\Delta] = \mathbf{Pr}[\xi_i = -\Delta] = \frac{1}{2}.$$

For each $m \leq n$, let $s = m\tau \in [0, t]$, and define

$$X_s = X_{m\tau} := \xi_1 + \dots + \xi_m.$$

Note that

$$X_{s+\tau} - X_s = \xi_{m+1}. \quad (1.19)$$

From (1.15) and (1.16) it is obvious that the expected value and the variance of X_s for $s = m\tau$ in this situation is given by

$$\mathbf{E}[X_s] = 0, \quad \text{Var}[X_s] = m\text{Var}(\xi_1) = \frac{s}{\tau}\Delta^2.$$

The idea is now to let $\tau \rightarrow 0$ and $\Delta \rightarrow 0$ in such a way that the variance of (X_s) remains constant. This will be guaranteed if τ and Δ are related by the equation

$$\Delta = \sqrt{\tau}, \quad (1.20)$$

which implies that for every $\tau > 0$ and every $s = m\tau$, with $0 \leq m \leq n$, we have

$$\mathbf{E}[X_s] = 0, \quad \text{Var}[X_s] = m\text{Var}(\xi_1) = s.$$

Note that the random processes constructed above are discrete-time processes, each defined for multiples of τ . Yet, as we let $\tau \rightarrow 0$, the points $m\tau$, $m = 0, 1, 2, \dots, n = t/\tau$ become increasingly denser in the interval $[0, t]$ and the random variable X_s converge in distribution to a random variable W_s which is called the *standard Wiener process* and is defined for all values of $0 \leq s \leq t$. Note that because for each τ the expected value of $X_{m\tau}$ is zero, this will also be inherited by the Wiener process, that is,

$$\mathbf{E}[W_s] = 0. \quad (1.21)$$

By a similar argument, the variance of W_s is the limit of the variances of respective discrete-time processes, hence

$$\text{Var}[W_s] = s.$$

The Wiener process is the building block for a large variety of continuous-time processes which are called *diffusion processes*. The standard Wiener process can thus be viewed as the simplest diffusion process.

1.17 The Distribution of the Wiener Process

Recall from the previous section that the Wiener process is a limit of discrete-time processes. In particular, for each $s = m\tau$, by (1.20)

$$X_s = \xi_1 + \dots + \xi_m,$$

where $\text{Var}[\xi_1] = \mathbf{E}[\xi^2] = \tau$. Various properties of the Wiener process can be deduced from here:

Given $s_0 < s_1 < \dots < s_r$ with $s_j = m_j\tau$ for $0 \leq j \leq r$, then

$$X_{s_1} - X_{s_0} = \xi_{s_0+1} + \dots + \xi_{s_1}$$

$$X_{s_2} - X_{s_1} = \xi_{s_1+1} + \dots + \xi_{s_2}$$

$$\dots \quad \dots$$

$$X_{s_r} - X_{s_{r-1}} = \xi_{s_{r-1}+1} + \dots + \xi_{s_r}.$$

Note that the increments $X_{s_1} - X_{s_0}, X_{s_2} - X_{s_1}, \dots, X_{s_r} - X_{s_{r-1}}$ are some *distinct* sets of independent random variables. This, in particular, implies that increments of the Wiener process over disjoint intervals are independent.

Another point is that for $s = m\tau$ and $t = n\tau$, with $n > m$, we have

$$X_t - X_s = \xi_{(m+1)\tau} + \dots + \xi_{n\tau}$$

as the sum of $n - m$ independent random variables. Note that

$$\mathbf{E}[X_t - X_s] = 0,$$

while

$$\text{Var}[X_t - X_s] = (n - m)\Delta^2 = t - s.$$

This shows that as $\tau \rightarrow 0$, $X_t - X_s$ converges to a normal random variable with variance $t - s$.

Using the central limit theorem, one can show that W_t is normally distributed with mean zero and variance t and its probability density function $p(x, t)$ is given by

$$p(x, t) = \frac{1}{\sqrt{(2\pi t)}} \exp\left(-\frac{x^2}{2t}\right).$$

In many applications we deal with random processes that have a tendency to increase or decrease. Surely such processes cannot be modeled using the Wiener process, which as (1.21) indicates, has a tendency to zero, which is called drift and will be defined next. In the following section we will study a modification of the Wiener process which makes it possible to use the Wiener process to construct such processes.

1.17.1 The Wiener Process with Drift: Two Approaches

The *Wiener process with drift* or generalized Wiener process is obtained by superimposing the Wiener process with a deterministic linear process. More precisely, for given constants μ and $\sigma > 0$, and an initial point z , we can define the process

$$X_t = z + \mu t + \sigma W_t. \quad (1.22)$$

Note that the deterministic term $z + \mu t$ has no effect on the variance, while the expected value will be moved in a linear way. From (1.22) we can easily see

$$\mathbf{E}[X_t] = \mu t, \quad \text{Var}[X_t] = \sigma^2 t.$$

Observe that the deterministic term μt does not influence the variance. Likewise, the constant σ only scales the noise W_t and does not affect the mean. It is also useful to remember that the Wiener process with drift X_t is a diffusion process which satisfies the stochastic differential equation given by

$$dX_t = \mu dt + \sigma dW_t. \quad (1.23)$$

Because X_t is obtained by a combination of scaling and translation of the standard Wiener process, it is easy to compute its density function, which is given by

$$p(x, t|z) = \frac{1}{\sigma \sqrt{2\pi t}} \exp\left(-\frac{(x - z - \mu t)^2}{2\sigma^2 t}\right).$$

Observe that X_t is normally distributed with expected value μt and variance $\sigma^2 t$; the parameter μ is called the drift coefficient or *drift rate* and the variance parameter σ^2 is called the *diffusion coefficient*.

One can also construct the Wiener process with drift as a limit of random walks on the integers with the difference that the random walks used in the process have to be biased. This fact is useful in the discretization process which is sometimes

used in applications. This can be done in at least two different ways. In the first approach, we set $s = m\tau$, where m is large and $\tau \approx 0$. We will divide the interval $[0, s]$ into m equal intervals of length $\tau = \frac{s}{m}$, and set

$$X_s = \xi_1 + \cdots + \xi_m.$$

As opposed to the previous case, here we will assume that ξ_i has the distribution given by

$$\Pr \left[\xi_i = +\Delta + \frac{\mu}{2}\tau \right] = \Pr \left[\xi_i = -\Delta + \frac{\mu}{2}\tau \right] = \frac{1}{2},$$

where $\Delta = \sigma\sqrt{\tau}$. It is easy to verify that $\mathbf{E}[\xi_i] = \mu\tau$ and

$$\text{Var}[\xi_i] = \mathbf{E}[\xi_i^2] - = \sigma^2\tau + \frac{\mu^2\tau^2}{4} \approx \sigma^2\tau$$

when $\tau \approx 0$. Setting

$$X_s = \xi_1 + \cdots + \xi_m$$

it follows that

$$\mathbf{E}[X_s] = \mu s, \quad \text{Var}[X_s] = s\sigma^2.$$

Various properties of the Wiener process can be deduced from here:

Given $s_0 < s_1 < \cdots < s_r$ with $s_j = m_j\tau$ for $0 \leq j \leq r$, then

$$\begin{aligned} X_{s_1} - X_{s_0} &= \zeta_{s_0+1} + \cdots + \zeta_{s_1} \\ X_{s_2} - X_{s_1} &= \zeta_{s_1+1} + \cdots + \zeta_{s_2} \\ &\dots &&\dots \\ X_{s_r} - X_{s_{r-1}} &= \zeta_{s_{r-1}+1} + \cdots + \zeta_{s_r}. \end{aligned}$$

Note that the increments $X_{s_1} - X_{s_0}, X_{s_2} - X_{s_1}, \dots, X_{s_r} - X_{s_{r-1}}$ are some *distinct* sets of independent random variables. This, in particular, implies that increments of the Wiener process with drift over disjoint intervals are independent.

Another point is that for $s = m\tau$ and $t = n\tau$, with $n > m$, we have

$$X_t - X_s = \zeta_{(m+1)\tau} + \cdots + \zeta_{n\tau}$$

which is the sum of $n - m$ independent random variables. Note that

$$\mathbf{E}[X_t - X_s] = (t - s)\mu,$$

while

$$\text{Var}[X_t - X_s] = (t - s)\sigma^2.$$

One can also construct the Wiener process with drift by changing the probabilities. This time, we consider a sequence of independent random variables ξ_1, ξ_2, \dots with the distribution given by

$$\Pr[\xi_i = \Delta] = \frac{1}{2} \left(1 + \mu \frac{\sqrt{\tau}}{\sigma} \right), \quad \Pr[\xi_i = -\Delta] = \frac{1}{2} \left(1 - \mu \frac{\sqrt{\tau}}{\sigma} \right), \tag{1.24}$$

where, as before, $\Delta = \sigma\sqrt{\tau}$, One can easily verify that

$$\mathbf{E}[\xi_i] = \mu\tau, \quad \text{Var}[\xi_i] = \sigma^2\tau - \mu^2\tau^2 \approx \sigma^2\tau$$

when $\tau \approx 0$. Note that as we increase m , and thereby decrease τ , we guarantee that the probabilities used in the definition of ξ_i are indeed between 0 and 1, and hence ξ_i are well-defined.

The drift rate μ plays a major role in modeling psychological processes and phenomena. The magnitude of the drift rate can reflect the quality of the stimuli. For instance, in a perceptual task the discriminability of two stimuli can be mapped onto μ , similar to d' in signal detection theory: the better they can be discriminated, the larger is μ . The drift rate can also be modeled as a function of stimulus properties. For instance, for preference choice situations, the drift rate may be determined by a sigmoid function as proposed in prospect theory (Kahneman & Tversky, 1979). The sign of the drift coefficient determines the direction the process takes, on average. With $\mu > 0$ the process moves up, on average, indicating more evidence for choosing one of the options. With $\mu < 0$ the process moves in the opposite direction, on average, in favor of choosing the alternative option. Setting $\sigma = 1$ and $\Delta = 1$, the transition probabilities in 1.14 can easily transform into a drift rate, and vice versa. For example, for $\mu = 0$, the probability is .5 of a step in either direction (simple random walk and standard Wiener process). With transition probabilities $p_{i,i+1} = .3$ for choosing A and $p_{i,i-1} = .7$ for choosing B from our introductory example, this yields $\mu = -.4$. That is, evidence mostly speaks for choosing option B.

1.17.2 Functional Relation between the Parameters

Consider a modeling problem that involves a Wiener process with drift given by

$$X_t = \mu t + \sigma W_t,$$

where W_t is the standard Wiener process. In many problems, it is desirable to reduce the number of involved parameters as much as possible, without reducing the generality of the situation.

In this section, we show how this can be done by explaining the functional relation between these parameters. In other words, we will see how these parameters vary under the change of temporal and spatial units.

The key role in the discussion is played by the *scale invariance* of the Wiener process. This means that if $\alpha > 0$ is a scaling factor, then the process

$$W'_t = \alpha^{-1} W_{\alpha^2 t}$$

is itself a standard Wiener process. This can be seen by checking the distribution of these random variables. For instance, we have

$$\text{Var}[W'_t] = \text{Var}\left[\alpha^{-1} W_{\alpha^2 t}\right] = \alpha^{-2}(\alpha^2 t) = \text{Var}[W_t].$$

This means that the Wiener process with drift $X_t = \mu t + \sigma W_t$ is equivalent to

$$X'_t = \mu t + \alpha^{-1} \sigma W_{\alpha^2 t} = \mu \alpha^{-2} (\alpha^2 t) + \sigma \alpha^{-1} W_{\alpha^2 t}.$$

Writing $\alpha^2 = \lambda$, we see that a rescaling of time by factor λ amounts to dividing μ by λ and σ by $\sqrt{\lambda}$. We will summarize this as follows:

$$t' = \lambda t, \quad \mu' = \lambda^{-1} \mu, \quad \sigma' = \lambda^{-1/2} \sigma.$$

Example 1.30 The process $X_t = \mu t + \sigma W_t$ is run until it hits the threshold C . We will show that in order to find a formula for the expected value of the reaction time $ER(\mu, \sigma, C)$, we can, without loss of generality, assume that $\sigma = 1$.

Denote by τ the stopping time defined by the smallest value of t for which $|X_t| = C$. Note that by the substitution $t' = \sigma^2 t$, we obtain an equivalent process $X'_t = (\mu/\sigma^2)t' + W_{t'}$, where the diffusion coefficient is 1. From here it follows that $ER(\mu, \sigma, C) = ER(\mu/\sigma^2, 1, C/\sigma^2)$. This function will be calculated in Example 1.36.

1.18 Computation with the Wiener Process

In this subsection, we will address a number of questions that are related to computations involving the Wiener process. Most of the formulas given in this section can also be found in Borodin and Salminen (2002). We will start with the Wiener process itself. Recall that for $t \geq 0$, the random variable W_t has normal distribution $N(0, t)$. Also, since $W_{s+t} = W_s + (W_{s+t} - W_s)$ is a sum of two independent normal random variables with variances s and t , respectively, we know the conditional density of W_{s+t} conditioned on W_s is given by

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-(y-x)^2/2t\right).$$

Note that this density does not depend on s , which is a reflection of the fact that the Wiener process is a Markov process. Recall that the distribution of the standard normal random variable is given by

$$\Phi(z) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx.$$

From here, one can see that the cumulative distribution function of W_t is given by $\Phi(z/\sqrt{t})$.

Example 1.31 Suppose W^x is the Wiener process starting at x . Compute that conditional probability $\Pr[W_t \leq y | W_0 = x]$.

Note that since $W_t - W_0$ is a normal random variable with variance t , we have

$$\Pr[W_t \leq y | W_0 = x] = \Phi\left(\frac{y-x}{\sqrt{t}}\right).$$

An important related process in decision theory is the running maximum of the Wiener process, which is described in the next paragraph.

The running maximum. Consider the maximum of the Wiener process over the time interval from 0 to t , and denote its value by M_t . In other words, $M_t = \max_{0 \leq s \leq t} W_s$. Note that since $W_0 = 0$, the running maximum M_t is always non-negative. Moreover, $W_t \leq M_t$ follows from the definition. One can show that the density function of M_t is given by

$$f_{M_t}(x) = \sqrt{\frac{2}{\pi t}} \exp\left(-\frac{x^2}{2t}\right), \quad x \geq 0.$$

In some applications, one requires the joint density of (M_t, W_t) . This is given by

$$f(x, y) = \frac{2(2x - y)}{\sqrt{2\pi t^{3/2}}} \exp\left(-\frac{(2x - y)^2}{2t}\right), \quad y \leq x, x \geq 0.$$

Example 1.32 For $a > 0$, let τ_z denote the stopping time defined by $\tau_z = \min\{t : W_t \geq z\}$. In other words, τ_z determines the first time that the Wiener process reaches the point z . Note that $\tau_z > t$ is equivalent to $M_t < z$. This implies that

$$\Pr[\tau_z > t] = \int_0^z f_{M_t}(x) dx.$$

Working out the above example, one can also compute the Laplace transform and the probability density function of τ_z . In the following theorem, the Wiener process is assumed to start at point x . The case of a standard Wiener process can be obtained by setting $x = 0$.

Theorem 1.33 Let τ_z be the stopping time defined in the previous paragraph. Then we have

1. The Laplace transform of τ_z is given by

$$\mathbf{E}[e^{-\alpha\tau_z}] = e^{-|x-z|\sqrt{2\alpha}}.$$

2. The density function of τ_z is given by

$$f_{\tau_z}(t) = \frac{|x-z|}{\sqrt{2\pi t^{3/2}}} \exp\left(-\frac{|x-z|^2}{2t}\right).$$

3. The probability that the Wiener process has not crossed another point y in the time between zero and τ_z is given by

$$\Pr\left[\sup_{0 \leq s \leq \tau_z} W_s < y\right] = \begin{cases} 1 & \text{if } x \leq z \\ \frac{y-x}{y-z} & z \leq x \leq y. \end{cases}$$

In the previous paragraph, the stopping time τ_z is defined using the Wiener process, and is hence dependent on it. In some applications, the Wiener process is stopped at a random time, *independently* of the process itself. A very useful special case is when this random time is determined by an exponential clock, running independently. More precisely, assume that W_t denotes a Wiener process without drift, and τ denotes a stopping time with exponential distribution with parameter λ and independent of W_t . The next theorem lists some of the properties of W_τ .

Theorem 1.34 For the stopping time τ defined in the previous paragraph, we have

1. $\mathbf{E}[e^{-\alpha\tau}] = \frac{\lambda}{\lambda+\alpha}.$
2. $\mathbf{E}[e^{i\beta W_\tau}] = \frac{2\lambda}{2\lambda+\beta^2} e^{i\beta x}.$
3. The density function of W_τ is given by

$$f_{W_\tau}(z) = \frac{\sqrt{\lambda}}{\sqrt{2}} \exp(-|z-x|\sqrt{2\lambda}).$$

Instead of deriving these formulas mathematically, we will show how one can derive such formulas by using a heuristic about the change in the Wiener process. Consider the Wiener process with drift given by $X_t = \mu t + \sigma W_t$. The change in the value of the process X_t from time t to $t + \Delta t$ can be expressed as

$$\Delta X = X_{t+\Delta t} - X_t = \mu \Delta t + \sigma dW_t.$$

From here, we can compute the average and variance of ΔX over this interval:

$$\begin{aligned}\mathbf{E}[\Delta X | X_t = x] &= \mu \Delta t + \sigma \mathbf{E}[\Delta W_t] = \mu \Delta t \\ \text{Var}[\Delta X | X_t = x] &= \sigma^2 \mathbf{E}[(\Delta W_t)^2] = \sigma^2 \Delta t,\end{aligned}$$

which implies that

$$\mathbf{E}[(\Delta X)^2 | X_t = x] = \sigma^2 \Delta t + \mu^2 \Delta t^2 \approx \sigma^2 \Delta t.$$

An application of this is shown in the next example.

Example 1.35 Consider the process $X_t = \mu t + \sigma W_t$, and let $a < x > b$. Let τ be the stopping time defined by the first time that the process X_t leaves the interval (a, b) . Hence, X_τ is either a or b . Then

$$\mathbf{Pr}[X_\tau = b | X_0 = x] = \frac{e^{-2\mu x/\sigma^2} - e^{-2\mu a/\sigma^2}}{e^{-2\mu b/\sigma^2} - e^{-2\mu a/\sigma^2}}.$$

In order to see this, let us denote the probability in question by $u(x)$ and view it as a function of x . As mentioned above, the argument given here is not mathematically accurate, but with some work can be turned into a rigorous argument. Consider the process at time $t + \Delta t$, and write $X_{t+\Delta t} = x + \Delta X$. Clearly the conditional probability that exiting the interval happens at b is now moved to $u(x + \Delta X)$. From here, we can write

$$u(x) = \mathbf{Pr}[X_\tau = b | X_0 = x] = \mathbf{E}[u(x + \Delta X)].$$

Now, we will use a Taylor expansion of the series for $u(x)$ by writing

$$u(x + \Delta X) \approx u(x) + u'(x)\Delta X + \frac{1}{2}u''(x)(\Delta X)^2.$$

By taking the expected value on both sides, we obtain

$$u(x) = u(x) + u'(x)\mathbf{E}[\Delta X] + \frac{1}{2}u''(x)\mathbf{E}[(\Delta X)^2].$$

Simplifying this equation yields

$$u(x) = u(x) + u'(x)\mu\Delta t + \frac{1}{2}u''(x)\sigma^2\Delta t,$$

which leads to the differential equation $\mu u'(x) + \frac{1}{2}\sigma^2 u''(x) = 0$. Solving this differential equation with the boundary conditions $u(a) = 0$ and $u(b) = 1$ yields the result.

Example 1.36 Let the process X_t and the stopping time τ be as in the previous example. Then

$$\mathbf{E}[\tau|X_0 = x] = \frac{1}{\mu}(u(x)(b-a) - (x-a))$$

where $u(x)$ is as in the previous example.

Let us denote the right-hand side by $v(x)$. As we move from time 0 to time Δt , the process moves from x to $x + \Delta X$. From here, we have

$$\begin{aligned} v(x)z &= \mathbf{E}[\tau|X_0 = x] = \mathbf{E}[\Delta t + \mathbf{E}[\tau - \Delta t|X_{\Delta t} = x + \Delta X]|X_0 = x] \\ &= \Delta t + \mathbf{E}[v(x + \Delta X)]. \end{aligned}$$

Again, using Taylor series development

$$v(x + \Delta X) \approx v(x) + v'(x)\Delta X + \frac{1}{2}v''(x)(\Delta X)^2$$

we obtain

$$v(x) \approx \Delta t + v(x) + \mu v(x)\Delta t + \frac{1}{2}v''(x)\sigma^2\Delta t.$$

Dividing by $\Delta t \rightarrow 0$ results in the differential equation

$$-1 = \mu v'(x) + \frac{1}{2}\sigma^2 v''(x).$$

Solving this together with the boundary condition $v(a) = v(b) = 0$ yields the result.

1.18.1 Simulation of the Wiener Process

When working with the Wiener process, it is often useful to be able to simulate a path of the process. One way of doing so is through approximation of the Wiener process with a random walk, when the step size of the walk is small. There are, however, other ways of simulating such a process. One possible approach that we will discuss here uses a sequence of independent standard normal random variables as an input and produces a path of the Wiener process as an output.

Suppose Z_0, Z_1, Z_2, \dots is a sequence of independent identically distributed random variables with standard normal distribution. For $0 \leq t \leq 1$, set

$$W_t = \frac{t}{\sqrt{\pi}} Z_0 + \sqrt{\frac{2}{\pi}} \sum_{m=1}^{\infty} \frac{\sin mt}{m} Z_m.$$

One can show that this infinite sum produces a random path of the Wiener process. In practice, we will work with a truncated finite version of this sum, namely

$$W_t^N = \frac{t}{\sqrt{\pi}} Z_0 + \sqrt{\frac{2}{\pi}} \sum_{m=1}^N \frac{\sin mt}{m} Z_m.$$

Note that by choosing M large enough one obtains increasingly better approximations.

This approximation works for values of $0 \leq t \leq 1$. When the Wiener process over an indefinite interval is needed we can use the fact that if W_t is a Wiener process for $0 \leq t \leq 1$, then

$$W'_t := (1+t)W_{\frac{1}{1+t}}$$

will define a Wiener process for all values of $t \geq 0$.

1.19 Multidimensional Wiener Process

Suppose W_t^1, \dots, W_t^n are standard independent Wiener processes. Independence means that for any $t_1 < \dots < t_m$, the vectors $(W^1(t_1), \dots, W^1(t_m))$ are independent vectors. Define, now, the vector-valued process

$$\mathbf{W}_t = (W_t^1, \dots, W_t^n).$$

This is called the n -dimensional Wiener process.

Recall that the chief interpretation of the Wiener process in psychology is information accrual. A decision-maker faced with two alternatives collects information concerning the choices, and the net amount of information for one (or the other) alternative can be described by a Wiener process with drift. When the number of alternatives is more than two, one can imagine “independent sources of information” which cannot be, strictly speaking, compared with each other. In other words, suppose there are d sources of evidence for a number of alternatives, but source i and source j are not directly comparable. In such cases, it will be useful to work with more than one Wiener process at the same time. We can do this by using a d -dimensional Wiener process. Before defining this in general, we will need to discuss the notion of a Gaussian vector.

1.20 Gaussian Vectors

Before we start discussing a Wiener process in higher dimensions, it is useful to define the notion of a Gaussian vector.

Definition 1.37 Let X_1, \dots, X_n be random variables. These random variables are said to be jointly Gaussian if any linear combination $c_1X_1 + \dots + c_nX_n$ has a Gaussian distribution.

Remark It is not always true that if X and Y are two Gaussian random variables then (X, Y) are jointly Gaussian. To see how this can fail, suppose X is a Gaussian random variable. Let Z be a signed Bernoulli random variable with parameter $p = \frac{1}{2}$, which is independent from X . Now, set $Y = XZ$. Note that, depending on whether $Z = 1$ or $Z = -1$, we have $Y = X$ and $Y = -X$. Independence of X and Y implies that Y itself is a Gaussian random variable.

On the other hand X, Y are not jointly Gaussian. To see this, note that

$$\Pr[X + Y = 0] = \Pr[Z = -1] = \frac{1}{2}.$$

Clearly, $X + Y$ is not a Gaussian random variable. On the other hand, the following is true:

Example 1.38 (Gaussian vectors) Suppose X_1, \dots, X_n are *independent* Gaussian random variables. Then they are jointly Gaussian. In order to see this, consider a linear combination $X = \sum_{j=1}^n c_jX_j$. In fact, a simple computation will show that X has distribution $N(\mu, \sigma^2)$, where

$$\mu = \sum_{j=1}^n c_j\mu_j, \quad \sigma^2 = \sum_{j=1}^n c_j^2\sigma_j^2.$$

1.20.1 The Covariance Matrix

Let us now assume that X_1, \dots, X_n are jointly Gaussian. Consider the matrix Σ defined by

$$\Sigma = (\text{cov}(X_i, X_j)).$$

This matrix is called the covariance matrix of X_1, \dots, X_n .

Example 1.39 Suppose X_1, \dots, X_n are independent Gaussian random variables and X_i has distribution $N(\mu_i, \sigma_i^2)$. Then it can be seen that the covariance matrix Σ is a diagonal matrix with diagonal entries $\sigma_1^2, \dots, \sigma_n^2$. In general, the off-diagonal entries of the covariance matrix contain information about how X_i and X_j correlate for $i \neq j$.

1.20.2 Invariance of Gaussian Vectors under Affine Transformations

In many applications, we would like to carry out linear transformations on Gaussian vectors. In this section, we will show that if Y is a Gaussian vector, then any affine transformation $X = AY + b$ of Y will also be a Gaussian vector.

Theorem 1.40 *Let Y be an $n \times 1$ Gaussian vector, and A an $m \times n$ matrix, and $b \in \mathbf{R}^m$. Then $X = AY + b$ is an $m \times 1$ Gaussian vector with*

$$\mu_X = A\mu_Y + b, \quad \Sigma_X = A\Sigma_Y A^\top.$$

In particular, if $\mu_Y = 0$ and $\Sigma_Y = I$, we have $\mu_X = b$ and $\Sigma_X = AA^\top$.

The converse of this theorem shows that every Gaussian vector is a linear combination of independent Gaussian random variables. We will first need the following lemma:

Theorem 1.41 *Let X be an $n \times 1$ Gaussian vector with Σ_Y invertible. Then there exists a Gaussian vector Y whose entries are independent, an invertible $n \times n$ matrix A , and a vector b such that $X = AY + b$.*

Proof Since Σ is an invertible positive-definite matrix, there exists a unique symmetric matrix A such that $AA^\top = \Sigma$. Let $Y = A^{-1}X$; applying the previous theorem, we have $\Sigma_Y = I$, which implies that the entries of X are independent. This proves the theorem. \square

Theorem 1.42 (Characteristic functions of Gaussian vectors) *Let $X = (X_1, \dots, X_n)^\top$ be a Gaussian vector with $\mathbf{E}[X] = \mu$ and $\text{cov}(X) = \Sigma$. The characteristic function of X is given by*

$$\phi_X(t) = \exp(it\langle t, \mu \rangle - \frac{1}{2}\langle t, \Sigma t \rangle^2). \quad (1.25)$$

Proof Applying Theorem 1.40 to the transpose of $t \in \mathbf{R}^n$, we see that $\langle t, X \rangle$ is a Gaussian random variable $N(\langle t, \mu \rangle, t^\top \Sigma t)$. \square

Theorem 1.43 *The density of a Gaussian vector is given by*

$$p_X(x_1, \dots, x_n) = (2\pi)^{-\frac{n}{2}} e^{-\frac{\|x\|^2}{2}}.$$

1.21 The Bessel Process

Another important process which appears in many computations involving the Wiener process is the Bessel process. Suppose $\mathbf{W}_t = (W_t^1, \dots, W_t^n)$ is an n -dimensional Wiener process. Define

$$R_t = \|\mathbf{W}_t\| = \left((W_t^1)^2 + \dots + (W_t^n)^2 \right)^{1/2}.$$

The process R_t defined in this way is called the *Bessel process* of order n . The following are some of the properties of the Bessel process.

1. R_t is a Markov process with continuous sample paths.
2. More precisely, the Bessel process is the solution to the following stochastic differential equation:

$$dR_t = dW_t + \frac{n-1}{2} \frac{dt}{R_t}.$$

Here, W_t is a standard Wiener process.

3. The transition density function from x to y is given by

$$p_t(x, y) = \frac{1}{t} \exp\left(-\frac{x^2 + y^2}{2t}\right) (xy)^{1-\frac{n}{2}} I_{\frac{n}{2}-1}(xy/y) y^{n-1}.$$

Here, $I_\nu(z)$ is the modified Bessel function defined by the power series

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+\nu}}{k! \Gamma(k+\nu+1)}.$$

1.22 The Ornstein–Uhlenbeck Process

The Wiener process provides a notable example of the family of Gaussian processes, in which random independent quantities are added, leading to a random continuous function. A broad class of such Gaussian processes that appear in pure and applied mathematics is called *diffusion processes*, the most prominent of which is the Wiener process we discussed earlier. In this section we will turn to another important diffusion process for cognitive science research, which is called the *Ornstein–Uhlenbeck process*, or sometimes, the OU process. It has properties that are interesting for model-building in psychology. For example, some properties account for assumptions concerning leakage or decay in information accumulation (Busemeyer & Townsend, 1992; Diederich, 1995; Smith, 1995; Usher & McClelland, 2001, 2004); primacy and recency effects in memory (Busemeyer & Townsend, 1993) or approach and avoidance conflicts in decision-making (Busemeyer & Townsend, 1993; Diederich, 2003).

The mathematical definition and proper derivation of diffusion processes are beyond the scope of this chapter; however, we will try to illustrate the idea behind the definition. Consider a stochastic process X_t , defined for positive values of t , with $X_0 = x$, and for a fixed time t , let ΔX denote the random change in the value of X from t to $t + \Delta t$, where Δt is a small time step. Hence, $\Delta X_t = X_{t+\Delta t} - X_t$ is a random variable whose mean and variance will depend on both t and Δt . Diffusions are processes for which both the average and variance of ΔX_t are linear functions of Δt as $\Delta t \rightarrow 0$. More precisely, diffusion processes are processes for which, conditioned on $X_t = x$, we can define the infinitesimal mean $\mu(x, t)$ and infinitesimal variance $\sigma^2(x, t)$ as follows:

- (a) The infinitesimal mean (or infinitesimal first moment or instantaneous mean or drift coefficient or drift rate) is defined as

$$\mu(x, t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{E} [\Delta X_t | X_t = x]}{\Delta t}. \quad (1.26)$$

- (b) The infinitesimal variance (or infinitesimal second moment or instantaneous variance or diffusion coefficient) is defined as

$$\sigma^2(x, t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Var} [\Delta X_t | X_t = x]}{\Delta t}. \quad (1.27)$$

Note that $\mu(x, t)$ and $\sigma^2(x, t)$ are deterministic functions of x and t . Once the drift and diffusion terms are specified, the transition probability density function can be determined and all required information about the process is provided.

The Wiener process is the simplest example of a diffusion process. We have already seen that for such a process $\mu(x, t) = \mu$ and $\sigma^2(x, t) = \sigma^2$, respectively. This can be seen by

$$\mu(x, t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{E} [\Delta X_t | X_t = x]}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\mu \Delta t}{\Delta t} = \mu$$

and

$$\sigma^2(x, t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{E} [(\Delta X_t^2 | X_t = x)]}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\sigma^2 \Delta t}{\Delta t} = \sigma^2.$$

We will now explain several ways of thinking about the OU process. However, before starting this discussion, we will give a simple example that contains some of the basic ingredients in the construction of the OU process.

1.22.1 The Ehrenfest Urn Model

Consider an urn containing n balls, some of which are blue and the rest red. At each time step, a random ball changes color. Denote by X_s the number of blue balls at time $s = 0, 1, \dots$. It is clear that X_s is a Markov process, whose transition probabilities are given by

$$p_{ij} = \begin{cases} \frac{i}{n} & \text{if } j = i - 1, \\ \frac{n-i}{n} & \text{if } j = i + 1. \end{cases}$$

Let us give a brief proof of this fact. Note that if $X_s = i$, then $X_{s+1} = i \pm 1$, depending on whether the random ball has color red or blue. Since the number of blue (red, respectively) balls is assumed to be i ($n-i$, respectively), the probability of transition to $j = i - 1$ ($j = i + 1$, respectively) is given by $p_{i,i-1} = \frac{i}{n}$ ($p_{i,i+1} = \frac{n-i}{n}$, respectively). Below is the transition matrix of this chain for $n = 5$:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 4/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 2/5 & 0 \\ 0 & 0 & 0 & 4/5 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Note, in particular, that the transition probabilities indicate reflecting boundaries. More precisely, when $i = n$, that is, when all the balls are blue, the next move

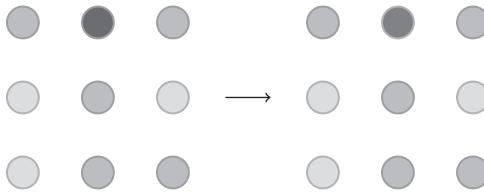


Figure 1.4 One step of the Markov chain in Ehrenfest's urn model.

takes the chain to $i = n - 1$ with probability 1, that is $p_{n,n-1} = 1$. Similarly, one can see that $p_{0,1} = 1$. More generally, if i is close to n , then the probability $p_{i,i-1} \simeq 1$. On the contrary, if i is small compared to n , then $p_{i,i+1} \simeq 1$. Roughly speaking, the probabilities are set up in a way to drive the chain toward a more balanced distribution of colors. This property is called *mean reversion*, and will also manifest itself in the OU process.

1.22.2 Construction of the Ornstein–Uhlenbeck Process

Let us now turn to the OU process. Historically, this process originated from an attempt to find a suitable model for the *velocity* of an infinitesimal grain acted upon by the molecules of a liquid in which it floated. In the applications in psychology, the velocity of the grain will be replaced by the velocity of the information arrival.

The most natural candidate is, of course, what we might obtain by simply differentiating the paths of the Wiener process, which, as explained before, are used to model the *position* of the grain. The problem is that one can show that the trajectories of the Wiener process are never differentiable. While we cannot give a proof of this fact here, we simply indicate that this relates to the fact that a Wiener process is expected to move by $\sqrt{\Delta t}$ over a period of Δt , and hence its slope $\sqrt{\Delta t}/\Delta t = 1/\sqrt{\Delta t}$ can get very large as $\Delta t \rightarrow 0$.

Instead of starting with the Wiener process, we will start with the assumption that over the period from t to $t + \Delta t$, the change in velocity of information accumulation is assumed to be influenced by two factors: the main underlying deterministic force is caused by specific tasks with their cognitive interpretation which slow down the motion of the process. For instance, in psychology the slowing down could be related to some information leakage (e.g., memory loss), decay in already accumulated evidence, or an avoidance tendency in conflicting decision situations. The analogous forces in physics are the frictional forces applied by the molecules of the medium in which the particle is moving.

The other factor is the random noise, over which there is no deterministic control. Again, this corresponds to unpredictable collisions that a particle undergoes in the medium. This will be naturally modeled by the Wiener process.

Assuming that the damping effect is directly proportional to the velocity leads to the equation

$$\mathbf{E} [\Delta X_t | X_t = x] = -\gamma x \Delta t + o(\Delta t), \quad (1.28)$$

where we have used the notation $\Delta X_t = X_{t+\Delta t} - X_t$.

The random noise contributes to the variance of the process, which will be proportional to Δt ; hence

$$\text{Var} [\Delta X_t | X_t = x] = \sigma^2 \Delta t + o(\Delta t). \quad (1.29)$$

In both of these equations $o(\Delta t)$ stands for an error whose size is of a smaller order than Δt .

Using Equation (1.26), one can see that the drift coefficient of the OU process described above is given by

$$\mu(x, t) = -\gamma x, \quad (1.30)$$

which is proportional to the deviation of the process from its mean at zero. In almost all applications, we assume that $\gamma > 0$. Then the negative coefficient $-\gamma$ has the effect that the process is mean-reverting: when $x > 0$, then the process has a negative drift pushing it toward zero; similarly, when $x < 0$, then the process will have a positive drift, again pushing it toward zero. Such a process is applicable to any model with a “leakage of information.”

On the other hand, for $\gamma < 0$, the process is pushed away from the mean at zero. This is applicable for a winner-takes-all situation, facilitation effects, or speeded responses in approach–conflict situations. However, the process might become unstable and, strictly speaking, it is not called the OU process for $\gamma < 0$.

Using (1.27), one can compute the diffusion coefficient of the OU process, which is given by

$$\sigma^2(x, t) = \sigma^2. \quad (1.31)$$

One can take (1.28) and (1.29) as the definition of the OU process. However, for computational purposes, this is not a useful definition. We will give the precise definition now.

Definition 1.44 The Ornstein–Uhlenbeck process with mean zero, drift coefficient γ and diffusion parameter σ is a Gaussian continuous Markov process X_t such that

1. The average at time t is given by

$$\mathbf{E}[X_t | X_0 = x] = \mu_t := xe^{-\gamma t}. \quad (1.32)$$

2. The variance is given by

$$\text{Var}[X_t | X_0 = x] = \sigma_t^2 := \frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma t}). \quad (1.33)$$

More generally, one can also consider OU processes where the mean is at an arbitrary point μ . Then, the value of the OU process at time t , X_t , is normally distributed with expected value

$$\mu_t = xe^{-\gamma t} + \mu (1 - e^{-\gamma t}). \quad (1.34)$$

The formula for the variance is the same. For the information of the reader, we will also point out that the OU process can also be defined as a solution of the stochastic differential equation

$$dX_t = \gamma(\mu - X_t)dt + \sigma dW_t.$$

We will not pursue this further in this note. The Ornstein–Uhlenbeck process can be constructed using the Wiener process. For instance, if W_t denotes the Wiener process, setting

$$X_t := xe^{-\gamma t} + \frac{\sigma e^{-\gamma t}}{\sqrt{2\gamma}} W_{\exp(2\gamma t)-1},$$

then one can see that X_t is an OU process. From here we see that the probability density function of X_t is given by

$$u(x, t) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(x-\mu_t)^2}{2\sigma_t^2}},$$

where μ_t and σ_t are given by Equations (1.34) and (1.33).

An interesting feature of the OU process is that as $t \rightarrow \infty$, its distribution converges to a stationary limit distribution. Note that as $t \rightarrow \infty$, we have $\mathbf{E}[X_t | X_0 = x] \rightarrow x$. In other words, the long-term average of the OU process is the initial point x . Similarly, one can check that the long-term variance of this process is $\sigma^2/2\gamma$. So, if an OU process has been running for a long time, one can expect its distribution to be close to a normal distribution with mean x , and variance $\sigma^2/2\gamma$.

1.22.3 The Ornstein–Uhlenbeck Process: From Discrete to Continuous

In this subsection we will study the relation between the Ehrenfest model and the OU process. Our exposition follows Karlin and Taylor (1981). Consider N balls, each of which is either blue or red. Each ball maintains its color for a random amount of time determined by an exponential random variable with parameter $\gamma > 0$. Suppose $\xi_i(t)$ is the random variable which takes values of $+1$ and -1 , determined by whether the ball i is blue or red at time t . This means that $\xi_i(t)$ is a two-state continuous-time process, whose transition probabilities are given by

$$\Pr[\xi_i(s+t) = \pm 1 | \xi_i(t) = 1] = \frac{1}{2} \pm \frac{1}{2} e^{-2\gamma t}.$$

In fact, using the notation of Section 1.15, one can view this as a Markov process with the matrix

$$\mathbf{G} = \begin{pmatrix} -\gamma & \gamma \\ \gamma & -\gamma \end{pmatrix}.$$

Let us also assume that initially each ball is blue or red with equal probabilities $1/2$. This implies that $\xi_i(0)$, and hence each $\xi_i(t)$ for all t have signed Bernoulli distribution with parameter $1/2$. A simple computation also shows that

$$\mathbf{E}[\xi_i(s)\xi_i(s+t)] = e^{-2\gamma s}.$$

Let us now consider the process defined by the difference between the number of blue and the number of red balls. In other words, we set

$$S_N(t) = \sum_{i=1}^N \xi_i(t).$$

By the central limit theorem, one expects the normalized sum $\frac{1}{\sqrt{N}}S_N(t)$ to be a Gaussian process with mean zero and covariance $\exp(-2\gamma|s-t|)$. One can check that the limit is an OU process.

1.22.3.1 OU Process as a Limit of a Birth–Death Chain

One can also construct the OU process as a limit of a *birth–death chain*. In general, the transition probabilities of a birth–death chain are given by

$$p_{ij} = \begin{cases} p_i & \text{if } j = i + 1 \\ q_i & \text{if } j = i - 1 \\ 1 - p_i - q_i & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

where $p_i, q_i \geq 0$ and $p_i + q_i \leq 0$. (p_i is called the birth rate, and q_i is called the death rate.)

Analogous to the Wiener process which was constructed as a limit of random walks, the OU process can be viewed as a limit of birth–death chains. The idea is to divide time to steps of size τ , and define

$$p_{ij} = \begin{cases} \frac{1}{2\alpha} \left(1 + (\mu - \gamma x_i) \frac{\sqrt{\tau}}{\sigma} \right) & \text{if } j = i + 1 \\ \frac{1}{2\alpha} \left(1 - (\mu - \gamma x_i) \frac{\sqrt{\tau}}{\sigma} \right) & \text{if } j = i - 1 \\ 1 - \frac{1}{\alpha} & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases} \quad (1.35)$$

The parameter, $\alpha > 1$, is a free parameter and is chosen to improve the approximation to a continuous-time process. It has no effect on choice probabilities.

1.23 Martingales and Their Applications

In this section, we will introduce the notion of a martingale, and state some of its properties. Martingales have proved to be indispensable tools in studying stochastic processes, and can also be used as powerful gadgets for carrying out calculations efficiently. Our emphasis in this section will be mostly on this aspect of the martingale theory.

Let (X_t) be a stochastic process, where $t \geq 0$ stands for discrete or continuous time. Let M_t be another process. We say that M_t is a martingale with respect to X_t if it satisfies the following properties:

1. The value of M_t can be determined by the values of X_s for $s \leq t$.
2. For any $t_2 > t_1 \geq 0$, we have

$$\mathbf{E}[M_t|X_s, s \leq t_1] = M_{t_1}.$$

3. For all $t > 0$, we have $\mathbf{E}[|M_t|]$ is finite.

The third condition is a technical one, and is expected to hold in all applications we have in mind. The first condition can be understood in the following way: although M_t is a random quantity, it can be deterministically computed based on the values of the underlying process X_s for all values of $s \leq t$. In particular, there is no need to “peek into the future” of the process X_s to determine M_t . The key condition in defining martingales is Condition 2, which basically states that given the information on X_s for $s \leq t$, the best guess for the value of M_t is M_{t_1} . This, in particular, implies that the average value of the martingale is constant as a function of t :

$$\mathbf{E}[M_t] = \mathbf{E}[M_0], \quad t \geq 0.$$

There are many interesting examples of martingales. Let us review some of them.

Example 1.45 A fair coin is tossed. Suppose $X_n = 1$ when the outcome of the n th toss is heads and $X_n = -1$ if the outcome of the n th toss is tails. Set $M_0 = 0$ and $M_n = X_1 + \dots + X_n$, for $n \geq 0$. Note that the value of M_n depends in a deterministic way on the values of X_1, \dots, X_n . Moreover, if $p < n$, then

$$\begin{aligned} \mathbf{E}[M_n|X_1, \dots, X_p] &= \mathbf{E}[M_p + X_{p+1} + \dots + X_n|X_1, \dots, X_p] \\ &= M_p + \sum_{j=p+1}^n \mathbf{E}[X_j|X_1, \dots, X_p] = M_p. \end{aligned} \tag{1.36}$$

The reason for the last equality is that for $j \geq p + 1$, X_j is independent of X_1, \dots, X_p , and hence the conditional expectation $\mathbf{E}[X_j|X_1, \dots, X_p]$ coincides with $\mathbf{E}[X_j] = 0$.

Example 1.46 Suppose X_1, X_2, \dots is a sequence of independent identically distributed random variables such that $\mathbf{E}[X_i] = 0$ and $\text{Var}[X_i] = \sigma^2$. We set $S_n = X_1 + \dots + X_n$ and define

$$M_n = S_n^2 - n\sigma^2.$$

One can easily check that M_n is a martingale.

Example 1.47 There are various examples of martingales that can be constructed using the Wiener process. In what follows, we will view W_t as the underlying process, and construct a variety of stochastic processes which are martingales with respect to W_t .

1. Suppose $M_t = \mu t + \sigma W_t$. It is easy to see that M_t is indeed a martingale with respect to W_t .

2. A less trivial example is $M_t = W_t^2 - t$. Note that

$$W_{t_2}^2 = (W_{t_1} + Z)^2 = W_{t_1}^2 + Z^2 + 2W_{t_1}Z,$$

where Z is a Gaussian random variable with variance $t-t_1$ which is independent of W_s for $s < t$. From here, we have

$$\begin{aligned}\mathbf{E}[M_t | W_s, s < t_1] &= \mathbf{E}\left[W_t^2 - t | W_s, s < t_1\right] \\ &= \mathbf{E}\left[W_{t_1}^2 + Z^2 + 2W_{t_1}Z - t | W_s, s < t_1\right] \quad (1.37) \\ &= W_{t_1}^2 t_1 + W_{t_1} \mathbf{E}[Z | W_s, s < t_1] \\ &= W_{t_1}^2 - t_1 = M_{t_1}.\end{aligned}$$

3. Another important example of martingales which is useful in many calculations is defined by

$$M_t = e^{aW_t - bt}.$$

1.23.1 Martingales and Stopping Times

Recall from above that the key defining property of the martingales is that the best forecast for the value of M_t given the information X_s for $s < t_1$ is M_{t_1} . Note that here, the forecast is made for the value at a deterministic point in the future. In many applications, we are interested in the expected value of M_t for a random future time. This, in particular, applies when this random time is a first passage time determined by the process X_t . It turns out that the answer is still positive in a large variety of cases. The main result in this case is the following theorem:

Theorem 1.48 (The optional sampling theorem) *Suppose that M_t is a martingale with respect to the underlying process X_s . Assume that τ is a stopping time with respect to the process X_s . Suppose further that there exists a deterministic time t_0 such that $\tau < t_0$, then we have*

$$\mathbf{E}[M_\tau] = \mathbf{E}[M_0].$$

Soon we will show the computational power of this theorem. However, before that, a cautionary comment is in order. The theorem above assumes the existence of t_0 such that $\tau < t_0$ almost surely. This assumption cannot be removed. To see this, consider the simple random walk on \mathbb{Z} , and let M_n denote the position of the random walk at time n . Hence, in particular, $M_0 = 0$. Let τ be the first time that the random walk visits 1. It can be easily seen that $\tau < \infty$. On the other hand, since $M_\tau = 1$, we have

$$\mathbf{E}[M_\tau] = 1 \neq 0 = \mathbf{E}[M_0].$$

This has to do with the fact that the stopping time τ is not bounded.

Let us show the power of the optional sampling theorem by two examples.

Example 1.49 Let us consider independent identically distributed random variables ξ_1, ξ_2, \dots , with $\mathbf{Pr}[\xi_i = 1] = \mathbf{Pr}[\xi_i = -1] = \frac{1}{2}$, and consider the random walk defined by $X_n = \xi_1 + \dots + \xi_n$. Suppose $a, b > 0$ are fixed integers, and the stopping time τ is defined to be the first time n such that $X_n = -a$ or $X_n = b$. Denote $Y = X_\tau$. It is clear that Y takes one of the two values $-a$ or b . Theorem 1.48 can be used to find the distribution of Y and also compute $\mathbf{E}[\tau]$.

Both computations are based on Theorem 1.48. Note that since X_n is the simple random walk on \mathbb{Z} , X_n itself is a martingale. Note that X_τ takes two values $-a$ and b . Let p_a and p_b denote the probabilities of these two events. Applying Theorem 1.48 to X_n we obtain

$$0 = \mathbf{E}[X_0] = \mathbf{E}[X_\tau] = (-a)p_a + bp_b.$$

Since $p_a + p_b = 1$, a simple computation shows that

$$p_a = \frac{b}{a+b}, p_b = \frac{a}{a+b}.$$

In order to compute $\mathbf{E}[\tau]$, we first note (Example 1.46) that $M_n := X_n^2 - n$ is a martingale. Applying Theorem 1.48 to M_n we obtain

$$0 = \mathbf{E}[M_\tau] = \mathbf{E}\left[X_\tau^2 - \tau\right] = a^2 p_a + b^2 p_b - \mathbf{E}[\tau].$$

From here, we obtain

$$\mathbf{E}[\tau] = a^2 p_b + b^2 p_a = ab.$$

Example 1.50 As before, let $X_t = \mu t + \sigma W_t$ be a Wiener process with drift, starting at zero. Let $a < 0 < b$, and define the stopping time τ by

$$\tau = \min\{t > 0 : X_t \geq b \text{ or } X_t \leq -a\}.$$

In other words, τ is the first time that the process X_t leaves the interval (a, b) . This, in particular, shows that X_τ is equal to a or b . We will use martingales to compute the probabilities

$$p_a = \mathbf{Pr}[X_\tau = a], \quad p_b = \mathbf{Pr}[X_\tau = b].$$

We dealt with this problem using a different method in Example 1.36.

Note that $p_a + p_b = 1$. Hence, we need one more linear equation involving p_a, p_b to determine the values of p_a, p_b . In order to do this, note that the process $M_t = e^{\gamma X_t}$ is a martingale. Using the above theorem, we have

$$\exp \gamma X_0 = \mathbf{E}[\exp(\gamma X_\tau)] = p_a \exp(\gamma a) + p_b \exp(\gamma b).$$

From here the values of p_a, p_b can be easily computed.

1.24 Decision-making with Multiple Attributes

The final section will show how we can use concepts and techniques discussed before to account for rather complex decision situations. Almost all sequential sampling models based on processes such as the Wiener process or the OU process assume a *single* integrated source for generating the evidence during the deliberation process leading to a decision. In particular, the integrated source may be based on multiple features or attributes, but all of these features or attributes are assumed to be combined and integrated into a single source of evidence, and this single source is used throughout the decision process until a final decision is reached. However, for many situations a sequential sampling model that represents evidence for the different process stages might be more appropriate than combining and integrating all information into a single source of evidence that drives the diffusion process. Diederich (1995, 1997) and Diederich and Oswald (2014) developed a generalization of the single-stage sequential model, assuming that each attribute of the stimulus arrangement is described by a separate sequential sampling process. In these works “attribute” is used in a very broad sense. It includes features or dimensions of the stimulus proper as well as information presented in different stages. Denote the set of attributes by κ . Since each attribute contains different information or appeals differently to the decision-maker, we allow for each attribute $k \in \kappa$ a process with different transition matrix \mathbf{P}_k on the same state space S . We will explain the ideas of this section by using the following example:

Example 1.51 A decision-maker deliberates buying a laptop. Two options are available, an expensive one with a high-resolution screen (call it C_1), and an inexpensive one with a low-resolution screen (call it C_2). The accumulated information by the decision-maker at any time will be represented by a state in the set $S = \{-2, -1, 0, 1, 2\}$. Here, state 2 represents settling for C_1 , and state -2 represents settling for C_2 . Other states, depending on their proximity to -2 and 2 , represent varying degrees of interest for one of these two options. This process can be formalized by associating a Markov chain to each attribute. More precisely, when price is considered by the decision-maker, the deliberation process is driven by a Markov chain with the state space S with the transition probabilities $p_{i,i+1} = 1/3$ and $p_{i,i-1} = 2/3$ with absorbing states ± 2 . Similarly, when screen resolution is considered, the deliberation process is driven by the Markov chain with the state space S and transition probabilities $p_{i,i+1} = 3/4$ and $p_{i,i-1} = 1/4$, with absorbing boundaries. We denote the set of attributes by $\kappa = \{1, 2\}$. Thus, we have the following transition matrix in the standard form:

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 2/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 2/3 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{R}_1 & \mathbf{Q}_1 \end{pmatrix}$$

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 3/4 & 0 \\ 0 & 0 & 1/4 & 0 & 3/4 \\ 0 & 3/4 & 0 & 1/4 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{R}_2 & \mathbf{Q}_2 \end{pmatrix}$$

Remark The procedure applies to the Wiener process and the OU process as well. For each of the $k \in \kappa$ attributes, a process with different drift and diffusion coefficients $\mu_k(x, t)$ and $\sigma_k^2(x, t)$ is defined. However, because the diffusion coefficient is merely a scaling parameter (see Section 1.17.2), we assume here the same coefficient for all attributes, i.e., $\sigma_k^2(x, t) = \sigma^2(x, t)$.

For the Wiener process we obtain drift rates μ_k for each of the $k \in K$ attributes resulting in transition probabilities (compare to (1.24))

$$p_{ij}^{(k)} = \begin{cases} \frac{1}{2} \left(1 + \mu_k \frac{\sqrt{\tau}}{\sigma} \right) & \text{if } j = i + 1 \\ \frac{1}{2} \left(1 - \mu_k \frac{\sqrt{\tau}}{\sigma} \right) & \text{if } j = i - 1 \\ 0 & \text{otherwise.} \end{cases} \quad (1.38)$$

Likewise for the Ornstein–Uhlenbeck process we assume different drift rate μ_k and decay parameter γ_k for the k th attribute, resulting in transition probabilities (compare to Equation (1.35))

$$p_{ij}^{(k)} = \begin{cases} \frac{1}{2\alpha} \left(1 + (\mu_k - \gamma_k x_i) \frac{\sqrt{\tau}}{\sigma} \right) & \text{if } j = i + 1 \\ \frac{1}{2\alpha} \left(1 - (\mu_k - \gamma_k x_i) \frac{\sqrt{\tau}}{\sigma} \right) & \text{if } j = i - 1 \\ 1 - \frac{1}{\alpha} & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases} \quad (1.39)$$

In principle, these constants (μ_k and γ_k) may also change within the sequence of attribute consideration. That is, when reconsidering an attribute, information evaluation may be different for the decision-maker compared to a previous consideration. Depending on the process we wish to model we can construct the matrices \mathbf{Q} and \mathbf{R} in (1.9) using (1.38) and/or (1.39). Note that we can also use both processes within a sequence of attributes.

In the following, we will see how these more complicated processes can also be modeled by building more complicated Markov chains/processes.

1.24.1 Time and Order Schedules

The information sampling is attribute-by-attribute, i.e., the finitely many attributes are considered one by one for a certain period of time in some order and possibly with repetition. The decision-maker switches attention from one attribute to the next during the time course of one trial. Here we distinguish different situations. The order in which attributes are processed may be known (e.g., fixed by the experimental design by providing pieces of information step by step). The time

each attribute is considered by the decision-maker may also be known (e.g., by displaying the information for a given time or by tracking the decision-maker's fixation time). Often, however, neither the processing order of attributes nor the point in time when the decision-maker switches attention from one attribute to the next one are known or can be inferred from the experimental set-up.

The specific order in which attributes are considered, called *order schedule*, as well as the times at which attention switches from one attribute to another one, called *time schedule*, is part of the model parameters. The order and time schedules may be given deterministically or randomly or a combination of both. When order and time schedules are deterministic we call the schedule *deterministic time and order schedule*. When both schedules are random, i.e., neither the sequence nor the point in time when the decision-maker switches attention from one to another attribute, then we call it a *random time and order schedule*. When one of the schedules is random and the other is deterministic we call it a *semi-random schedule*. For instance, the order in which attributes are known is known, but the exact time when the decision-maker switches from one attribute to another is not known.

Formally, we assume that attention switches from one attribute to the next in a sequence of *attention-switching times*.

$$T_0 = T_{start} = 0 < T_1 < T_2 < \dots < T_L = T_{end}, \quad (1.40)$$

with T_{end} representing the maximum duration of the decision process, called *time horizon*. On a theoretical level, it is possible to assume $T_{end} = \infty$ (infinite time horizon) and $L = \infty$ (infinite attention switching). We denote by $\Delta T_l = (T_{l-1}, T_l]$ the l th *attention time interval*. That is the time the decision-maker spends on considering an attribute. Note that T may be a constant (deterministic time schedule) or a random variable (random time schedule).

Definition 1.52 A time and order schedule consists of a sequence $\{T_l\}_{l=1,\dots,L}$ of attention-switching times, and a sequence $\{k_l \in \{1, \dots, \kappa\}\}_{l=1,\dots,L}$ of attribute indices which specifies that during the attention time interval ΔT_l the k_l th attribute is considered. At attention-switching time T_l , $l = 1, \dots, L - 1$, attention switches from attribute k_l to attribute k_{l+1} .

The process X_t determined by such a schedule is a *piecewise* diffusion process with fixed parameters in each interval ΔT_l .

We start with the deterministic time and order schedule.

Example 1.53 (continued)

In order to model the situation in Example 1.52 using a Markov chain, we will extend the state space.

Suppose attributes are to be considered in the order k_1, \dots, k_l . For now, we will assume that each attribute will be considered only for one unit of time. Define the extended state space by

$$\tilde{S} = \{(s, k) : s \in S, k \in K\}.$$

In other words, the extended state space not only keeps track of the state in which the chain is, but also of the attribute which is under consideration at the time. In the example above, we have \tilde{S} which contains 10 elements of the form (i, k) with $-2 \leq i \leq 2$ and $k = 1, 2$.

Let us first consider the case that attention switches from one attribute to the next in a fixed sequence. For instance, let us consider the case that the decision-maker spends one unit of time on the first attribute and one unit of time on the second attribute, and repeats this. For instance, if the decision-maker is at point 0, and is deliberating price, then in the next step they will be at 1 with probability $1/3$ and at -1 with probability $2/3$ and shifts their attention to screen resolution. We order the states of the chain as

$$\tilde{s}_1 = (-2, 1), \tilde{s}_2 = (-1, 1), \dots, \tilde{s}_5 = (+2, 1), \tilde{s}_6 = (-2, 2), \dots, \tilde{s}_{10} = (+2, 2).$$

It is easy to see that the transition matrix of the extended Markov chain is the 10×10 matrix given by the block matrix

$$\mathbf{P} = \begin{pmatrix} 0 & \mathbf{P}_1 \\ \mathbf{P}_2 & 0 \end{pmatrix}.$$

Assume that the decision-maker spends two units of time deliberating price and one unit of time deliberating the screen resolution. In this case, we will need a larger state space that distinguishes between the first and the second round of deliberating price. When done so, the transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} 0 & \mathbf{P}_1 & 0 \\ 0 & 0 & \mathbf{P}_1 \\ \mathbf{P}_2 & 0 & 0 \end{pmatrix}.$$

1.24.1.1 Deterministic Time and Order Schedule

We will now derive the choice probability, first passage time distribution and mean response time for choosing one of the alternatives, here A , for a deterministic time and order schedule.

The evidence-accumulation process for attribute k_1 , which is considered first, evolves until time t_1 when the second attribute k_2 comes into consideration, triggering a change in the accumulation process. This attribute in turn is considered until time t_2 when a third attribute k_3 is considered and so forth until a decision is initiated (or t_L is reached). The probability p_A to choose alternative A (that is, the process is absorbed at state a (Section 1.10), i.e., the criterion θ_A), is therefore

$$\begin{aligned} p_A = \mathbf{Pr}[T < t_1 \cap \text{choose } A] + \mathbf{Pr}[t_1 \leq T < t_2 \cap \text{choose } A] + \dots \\ + \mathbf{Pr}[t_{l-1} \leq T < t_l \cap \text{choose } A] + \dots + \mathbf{Pr}[t_L \leq T \cap \text{choose } A]. \end{aligned}$$

With the switching times t_l replaced by integers $n_l \approx t_l/\tau$, and invoking Equations (1.10) and (1.11), the choice probability for choosing option A is

$$p_A = \mathbf{Z}' \sum_{i=1}^{n_1} \mathbf{Q}_{k_1}^{i-1} \mathbf{R}_{A,k_1} + \mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \sum_{i=n_1+1}^{n_2} \mathbf{Q}_{k_2}^{i-(n_1+1)} \mathbf{R}_{A,k_2} + \dots \\ \dots + \mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \dots \mathbf{Q}_{k_{L-1}}^{n_{L-1}-n_{L-2}} \sum_{i=n_{L-1}+1}^{n_L} \mathbf{Q}_{k_L}^{i-(n_{L-1}+1)} \mathbf{R}_{A,k_L}. \quad (1.41)$$

The evidence-accumulation process for a successive attribute starts with the final evidence state of the previous attribute. Note that $\mathbf{Z}' \mathbf{Q}_{k_1}^{n_1}$ to $\mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \dots \mathbf{Q}_{k_{L-1}}^{n_{L-1}-n_{L-2}}$ are defective distributions, i.e., the entries of these vectors do not sum to 1, for the states of the random walk at discrete times n_1, \dots, n_{L-1} . Further, note that the stochastic process is time-homogeneous within each time interval $[0, t_1]$ to $[t_{L-1}, t_L]$ but non-homogeneous across $[0, t_L]$ (see Diederich, 1995).

The first passage time distribution given option A has been chosen is

$$\Pr(T = n|A) = \frac{1}{p_A} \begin{cases} \mathbf{Z}' \mathbf{Q}_{k_1}^n \mathbf{R}_{A,k_1} & 0 = n_0 < n \leq n_1 \\ \mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \mathbf{Q}_{k_2}^{n-n_1} \mathbf{R}_{A,k_2} & n_1 < n \leq n_2, \\ \dots \\ \mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \dots \mathbf{Q}_{k_{L-1}}^{n_{L-1}-n_{L-2}} \mathbf{Q}_{k_L}^{n-n_{L-1}} \mathbf{R}_{A,k_L} & n_{L-1} < n < \infty. \end{cases} \quad (1.42)$$

Similarly, the mean response time for choosing alternative A is approximated as

$$ET_A = \frac{\tau}{p_A} \left[\mathbf{Z}' \sum_{i=1}^{n_1} i \mathbf{Q}_{k_1}^{i-1} \mathbf{R}_{A,k_1} + \mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \sum_{i=n_1+1}^{n_2} i \mathbf{Q}_{k_2}^{i-(n_1+1)} \mathbf{R}_{A,k_2} + \dots \right. \\ \left. \dots + \mathbf{Z}' \mathbf{Q}_{k_1}^{n_1} \dots \mathbf{Q}_{k_{L-1}}^{n_{L-1}-n_{L-2}} \sum_{i=n_{L-1}+1}^{n_L} i \mathbf{Q}_{k_L}^{i-(n_{L-1}+1)} \mathbf{R}_{A,k_L} \right]. \quad (1.43)$$

The probability and the mean response time for choosing alternative B can be determined similarly. Note that $p_0 := 1 - (p_A + p_B)$, the probability of not making a decision until the final time t_L , is strictly positive if $t_L \approx n_L \tau < \infty$.

1.24.1.2 Random Time and Order Schedule

The above derivation of formulas for choice probabilities and mean response times for a deterministic time and order schedule have counterparts for random schedules which we describe next in three steps. But first we continue with our example.

Example 1.54 (continued)

Now, let us assume that the decision-maker switches attention from one attribute to another in a random way. We will also assume that attention-switching itself is a Markov process with the transition matrix given by the matrix

$$\mathbf{J} = \begin{pmatrix} 1/5 & 4/5 \\ 2/3 & 1/3 \end{pmatrix}.$$

For instance, if the price is the attribute under consideration at a given step, in the next round, it will still be the attribute under consideration with probability $1/5$. With probability $4/5$ the attention will be switched to resolution. Again, let us consider the extended state space \widetilde{S} . Some of the transition probabilities can be easily calculated:

$$P(0,1),(1,1) = \frac{1}{5} \cdot \frac{1}{3}, \quad P(0,1),(-1,1) = \frac{1}{5} \cdot \frac{2}{3}, \quad P(0,1),(1,2) = \frac{4}{5} \cdot \frac{1}{3}, \quad P(0,1),(-1,2) = \frac{4}{5} \cdot \frac{2}{3}.$$

On the other hand, note that $(\pm 2, 1)$ and $(\pm 2, 2)$ are the absorbing states of this chain. From here, we can easily see that the transition matrix of this chain is given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 \\ \mathbf{R}_1 & 0 & \frac{1}{5}\mathbf{Q}_1 & \frac{4}{5}\mathbf{Q}_1 \\ 0 & \mathbf{R}_2 & \frac{2}{3}\mathbf{Q}_2 & \frac{1}{3}\mathbf{Q}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{R} & \mathbf{Q} \end{pmatrix}$$

Note that this is the standard form of the matrix \mathbf{P} with

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & 0 \\ 0 & \mathbf{R}_2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \frac{1}{5}\mathbf{Q}_1 & \frac{4}{5}\mathbf{Q}_1 \\ \frac{2}{3}\mathbf{Q}_2 & \frac{1}{3}\mathbf{Q}_2 \end{pmatrix}.$$

Remark Analogous computations can be carried out in general. Suppose that for each of the $k = \{1, 2, \dots, K\}$ attributes we have a stochastic matrix \mathbf{P}_k defining a Markov chain. Write \mathbf{P}_k in the standard form

$$\mathbf{P}_k = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{R}_k & \mathbf{Q}_k \end{pmatrix}.$$

Suppose transition between the attributes will take place with another stochastic matrix \mathbf{J} . Then, we can model the extended Markov chain with the state space $S \times \kappa$, and the transition matrix of the new chain will be given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{R} & \mathbf{Q} \end{pmatrix}$$

where \mathbf{R} is the matrix with $\mathbf{R}_1, \dots, \mathbf{R}_k$ on its diagonal, and \mathbf{Q} is given by the Kronecker product

$$\mathbf{Q} = \mathbf{J} \otimes \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_1 & \dots & \mathbf{Q}_1 \\ \mathbf{Q}_2 & \mathbf{Q}_2 & \dots & \mathbf{Q}_2 \\ \vdots & \vdots & & \vdots \\ \mathbf{Q}_k & \mathbf{Q}_k & \dots & \mathbf{Q}_k \end{pmatrix}.$$

We now continue to describe the random schedules in detail.

Random order schedule. For generating the attribute order $\{k_l\}_{l=1,\dots,L}$, we consider stochastic $\kappa \times \kappa$ matrices $\mathbf{D}^{(l)}$ such that $d_{k'k}^{(l)} \geq 0$ describes the probability

with which attention switches from the k' th attribute to the k th attribute at switching time $t_l \approx \tau n_l$, $l = 1, \dots, L - 1$. To avoid no attention-switching between attributes we can set $d_{kk}^{(l)} = 0$.

For example, for two attributes, $\kappa = 2$, this results in $d_{11}^{(l)} = d_{22}^{(l)} = 0$, $d_{12}^{(l)} = d_{21}^{(l)} = 1$, or, written as a stochastic matrix, in

$$\mathbf{D}^{(l)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (1.44)$$

The attribute sequence is either $(1, 2, 1, 2, \dots)$ or $(2, 1, 2, 1, \dots)$, depending on whether $k_1 = 1$ or $k_1 = 2$.

The next example involves three attributes, $\kappa = 3$, three attention time intervals, i.e., a time schedule of length $L = 3$, and an unbiased preference for choosing any of the two remaining attributes at attention-switching time t_1 and some other (biased) probabilities to switch from one attribute to the next at attention-switching time T_2 . The stochastic matrices $\mathbf{D}^{(l)}$ for this situation are

$$\mathbf{D}^{(1)} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \text{ and, for instance, } \mathbf{D}^{(2)} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 3/4 & 1/4 & 0 \end{bmatrix}.$$

For attribute $k_1 = 1$ there are four possible order sequences, $(1, 2, 1)$, $(1, 3, 1)$, $(1, 2, 3)$ and $(1, 3, 2)$. The sequence $(1, 2, 1)$ would occur with probability $d_{12}^{(1)}$. $d_{21}^{(2)} = 1/2 \cdot 1 = 1/2$. The probabilities for the remaining sequences are $3/8$, 0 , and $1/8$, respectively. For attributes $k_2 = 1$ and $k_3 = 1$, the order sequences and probabilities of occurrence can be determined accordingly.

Random time schedule. For generating a random time schedule we assume that the attention time interval ΔT , i.e., the number of discrete time steps during which attention is paid to the k th attribute, is a discrete random variable with given distribution(s). In principle, it allows for different distributions for each attribute and attribute order $\{k_l\}_{l=1,\dots,L}$ to model, for instance, time pressure and other temporal effects. However, here we assume one and the same distribution type for attention times across all attributes, and allow for different parameters only.

For instance, the *geometric distribution* (as implicitly considered in Diederich, 1997) is given by

$$\Pr(\Delta T = n) = r(1 - r)^{n-1}, \quad n = 1, 2, \dots,$$

and characterized by a single parameter $r > 0$, with expected value $\mathbf{E}(\Delta T) = 1/r$ and variance $\text{Var}(\Delta T) = (1 - r)/r^2$; the *Poisson distribution* given by

$$\Pr(\Delta T = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n = 0, 1, 2, \dots,$$

characterized by a single parameter $\lambda > 0$, with expectation λ and variance λ ; the *binomial distribution* given by

$$\Pr(\Delta T = n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}, \quad n = 0, 1, \dots, N,$$

with parameters N and p , with expectation Np and variance $Np(1 - p)$; and the *uniform distribution* is defined as

$$\Pr(\Delta T = n) = \frac{1}{2M + 1}, \quad n = N - M, \dots, N + M,$$

with parameters N and $M = 0, 1, \dots, N-1$ and expectation N and variance $M(M + 1)/3$.

1.24.2 Implementation of Random Schedules

To see how the random attention time is implemented into the piecewise stochastic process – the multi-stage model – and to understand the recursive computation of choice probabilities and response time order schedules of length $L = l$, we first consider the special cases $L = 1$ and $L = 2$, and illustrate the derivation of some distribution types of the random variable ΔT generating attention times by providing concrete formulas. Derivations are based on Diederich and Oswald (2014).

In general, the distribution for ΔT is given by its probability mass distribution function (pdf) and cumulative distribution function (cdf)

$$\Pr(\Delta T = n) = p_{n,k}, \quad \Pr(\Delta T \leq n) = f_{n,k} := \sum_{i=0}^n p_{i,k}, \quad n = 0, 1, \dots \quad (1.45)$$

1.24.2.1 Order Schedule of Length $L = 1$

We start with an order schedule of length $L = 1$ and will drop the index l from the notation introduced in the previous subsection. Because the probability of choosing alternative A at the i th step is given by $\mathbf{Z}'\mathbf{Q}_k^{i-1}\mathbf{R}_{A,k}$, $i = 1, \dots, \Delta T$, and ΔT is a random variable distributed according to (1.45) we get

$$\begin{aligned} p_{A,k} &= \sum_{n=1}^{\infty} p_{n,k} \mathbf{Z}' \left(\sum_{i=1}^n \mathbf{Q}_k^{i-1} \right) \mathbf{R}_{A,k} \\ &= \mathbf{Z}' \left[\sum_{i=1}^{\infty} \left(\sum_{n=i}^{\infty} p_{n,k} \right) \mathbf{Q}_k^{i-1} \right] \mathbf{R}_{A,k} \\ &= \mathbf{Z}' \left[\sum_{i=0}^{\infty} (1 - f_{i,k}) \mathbf{Q}_k^i \right] \mathbf{R}_{A,k}. \end{aligned} \quad (1.46)$$

A similar formula holds for $p_{B,k}$ (for alternative B). Note that generally $p_{A,k} + p_{B,k} < 1$ because $(1 - f_k)\mathbf{Q}_k$ is no longer a transition probability matrix, i.e., the rows do not add up to 1.

To avoid repetition, we introduce the row vector $p_{AB,k} := [p_{B,k}, p_{A,k}]$, then

$$p_{AB,k} = \mathbf{Z}'\mathbf{V}_k, \quad \mathbf{V}_k := \left[\sum_{i=0}^{\infty} (1 - f_{i,k}) \mathbf{Q}_k^i \right] \mathbf{R}_k. \quad (1.47)$$

The $2 \times (m - 2)$ matrix \mathbf{V}_k depends on the parameters describing the process for attribute k ((1.38) or (1.39) build \mathbf{Q}_k , \mathbf{R}_k), and on the chosen attention time distribution and the cdf ($f_{n,k}$).

For specific attention time distributions discussed earlier these matrices may be precomputed. In some cases closed-form expressions can be found. For example, for the geometric distribution with parameter $r = r_k$, the matrix \mathbf{V}_k becomes

$$\begin{aligned}\mathbf{V}_k &= \sum_{i=0}^{\infty} \left(\sum_{j=i+1}^{\infty} r_k(1-r_k)^{j-1} \right) \mathbf{Q}_k^i \mathbf{R}_k \\ &= \sum_{i=0}^{\infty} (1-r_k)^i \mathbf{Q}_k^i \mathbf{R}_k = (\mathbf{I} - (1-r_k)\mathbf{Q}_k)^{-1} \mathbf{R}_k,\end{aligned}\quad (1.48)$$

which is a closed-form expression using a tridiagonal matrix inversion.

For the uniformly distributed attention time the cumulative distribution function is

$$f_{i,k} = \frac{i - N + M + 1}{2M + 1} \text{ and } (1 - f_{i,k}) = \frac{M + N - i}{2M + 1}$$

(the survivor function). Therefore, matrix V_k is

$$\mathbf{V}_k = \left(\sum_{i=0}^{N-M-1} \mathbf{Q}_k^i + \sum_{i=N-M}^{N+M-1} \frac{N+M-i}{2M+1} \mathbf{Q}_k^i \right) \mathbf{R}_k. \quad (1.49)$$

We can derive formulas for mean response times in a similar way. Let $\mathbf{ET}_{A,k}$ denote the mean response time for reaching alternative A when considering the k th attribute for a random time ΔT distributed according to (1.45). Then $\mathbf{ET}_{A,k} \approx \tau et_{A,k}/p_{A,k}$, where

$$\begin{aligned}et_{A,k} &= \sum_{n=1}^{\infty} p_{n,k} \left(\sum_{i=0}^{n-1} (i+1) \mathbf{Z}' \mathbf{Q}_k^i \right) \mathbf{R}_{A,k} \\ &= \mathbf{Z}' \left[\sum_{i=0}^{\infty} \left(\sum_{n=i+1}^{\infty} p_{n,k} \right) (i+1) \mathbf{Q}_k^i \right] \mathbf{R}_{A,k} \\ &= \mathbf{Z}' \left[\sum_{i=0}^{\infty} (1-f_{i,k})(i+1) \mathbf{Q}_k^i \right] \mathbf{R}_{A,k}.\end{aligned}\quad (1.50)$$

The same holds for $\mathbf{ET}_{B,k}$ and $et_{B,k}$. Thus, similar to (1.47), we write

$$et_{AB,k} := [et_{B,k}, et_{A,k}] = \mathbf{Z}' \mathbf{W}_k, \quad \mathbf{W}_k := \left[\sum_{i=0}^{\infty} (1-f_{i,k})(i+1) \mathbf{Q}_k^i \right] \mathbf{R}_k \quad (1.51)$$

for $k = 1, \dots, K$ attributes. In the computer programs the matrices \mathbf{V}_k and \mathbf{W}_k can be precomputed to any accuracy.

1.24.2.2 Order Schedule of Length $L = 2$

We next consider the $L = 2$ case. For simplicity we assume that the attention time distribution is the same for all attributes. To save on indices, denote $k_1 \equiv k'$, $k_2 \equiv k$, and $\mathbf{D}^{(1)} \equiv \mathbf{D}$ (the matrix in (1.44), responsible for the random choice of k given any k'). The choice probability vector $p_{AB,k',k}$ for reaching the criterion for choosing alternatives A or B with attribute order (k', k) consists of two parts:

- (1) the probabilities of reaching a boundary while considering the k' th attribute, which is processed first (i.e., $T_A/\tau \leq \Delta T'$, where $\Delta T'$ is the randomly generated attention time for the first attribute k') plus
- (2) the probabilities that the process reaches a boundary when considering the second attribute (i.e., $\tau \Delta T' < T_A/\tau \leq \Delta T' + \Delta T$, where ΔT is the randomly (and independently) generated attention time for the second attribute k). Furthermore, k itself is randomly chosen according to the entries in the k' th row of \mathbf{D} . Therefore, according to (1.47), for each fixed $k_1 = k'$ and $n_1 = \Delta T'$, the probabilities for initiating a choice response after time n_1 are given by

$$\begin{aligned} & \left[\Pr(\Delta T' < \frac{T_B}{\tau} < \infty), \Pr(\Delta T' < \frac{T_A}{\tau} < \infty) \right]_{n_1=\Delta T', k_1=k'} \\ & \approx \sum_{k=1}^K d_{k'k} \mathbf{Z}' \mathbf{Q}_{k'}^{n_1} \mathbf{V}_k = \mathbf{Z}' \mathbf{Q}_{k'}^{\Delta T'} \left(\sum_{k=1}^K d_{k'k} \mathbf{V}_k \right). \end{aligned}$$

For $L = 2$, the choice probabilities, with $k_1 = k'$ fixed, are therefore

$$\begin{aligned} [p_B, p_A]_{k_1=k'} &= \mathbf{Z}' \mathbf{V}_{k'} + \sum_{n \geq 0} p_{n,k'} \mathbf{Z}' \mathbf{Q}_{k'}^n \left(\sum_{k=1}^K d_{k'k} \mathbf{V}_k \right) \\ &= \mathbf{Z}' \left[\mathbf{V}_{k'} + \left(\sum_{n \geq 0} p_{n,k'} \mathbf{Q}_{k'}^n \right) \left(\sum_{m=1}^M d_{k'm} \mathbf{V}_m \right) \right] \\ &= \mathbf{Z}' \left[\mathbf{V}_{k'} + \mathbf{B}_{k'} \left(\sum_{k=1}^K d_{k'k} \mathbf{V}_k \right) \right], \quad k' = 1, \dots, K, \quad (1.52) \end{aligned}$$

where

$$\mathbf{B}_k := \sum_{n \geq 0} p_{n,k} \mathbf{Q}_k^n, \quad k = 1, \dots, K, \quad (1.53)$$

are $(m - 2) \times (m - 2)$ matrices depending on the parameters for attribute k and the type of attention time distribution.

By using similar notation and arguments as for choice probabilities, the quantities $et_{A,k',k}$, $et_{B,k',k}$ have a part before and after $\Delta T'$. This, together with (1.50) and (1.51), gives

$$\begin{aligned}
et_{AB}|_{k_1=k'} &= \mathbf{Z}' \mathbf{W}_{k'} + \sum_{n=0}^{\infty} p_{n,k} \mathbf{Z}' \mathbf{Q}_{k'}^n \left(\sum_{k=1}^K d_{k'k} (n \mathbf{V}_k + \mathbf{W}_k) \right) \\
&= \mathbf{Z}' \left[\mathbf{W}_{k'} + \left(\sum_{i=0}^{\infty} p_{i,k'} i \mathbf{Q}_{k'}^i \right) \left(\sum_{k=1}^K d_{k'k} \mathbf{V}_k \right) \right. \\
&\quad \left. + \left(\sum_{i=0}^{\infty} p_{i,k'} \mathbf{Q}_{k'}^i \right) \left(\sum_{k=1}^K d_{k'k} \mathbf{W}_k \right) \right] \\
&= \mathbf{Z}' \left[\mathbf{W}_{k'} + \mathbf{C}_{k'} \left(\sum_{k=1}^K d_{k'k} \mathbf{V}_k \right) + \mathbf{B}_{k'} \left(\sum_{k=1}^K d_{k'k} \mathbf{W}_k \right) \right],
\end{aligned}$$

where

$$\mathbf{C}_k = \sum_{n \geq 0} p_{n,k} n \mathbf{Q}_k^n, \quad k = 1, \dots, K. \quad (1.54)$$

1.24.2.3 Order Schedule of Arbitrary Length L

For arbitrary L , it is more convenient to write the resulting recursion in terms of *block-matrix–vector operations*. Denote by

- Z** the $\kappa \times 1$ array with each entry equal to the initial distribution \mathbf{Z} for each of the K attributes (think of \mathbf{Z}' as its transpose, a $1 \times K$ array with entries \mathbf{Z}').
- B** the $K \times K$ diagonal array with the \mathbf{B}_k on the diagonal (see 1.53). Here we assume that each attribute has the same attention time distribution (with different parameters). Note, however, that it is possible to assume different distribution for different attributes.
- C** the $K \times K$ diagonal array with the \mathbf{C}_k on the diagonal (see 1.54) with the same attention time distributions corresponding to **B**.
- I** the $K \times K$ diagonal array, with identity matrices **I** of the appropriate size on the diagonal.
- V** the $K \times 1$ array with the \mathbf{V}_k as entries according to (1.47).
- W** the $K \times 1$ array with the \mathbf{W}_k as entries according to (1.51).

Using this notation the choice probability for time schedule of length $L = 2$ can be written as

$$\begin{aligned}
\mathbf{p}_{AB} &= \mathbf{Z}' (\mathbf{I} + \mathbf{BD}) \mathbf{V} \\
&= [\mathbf{Z}'_1 \mathbf{Z}'_2] \left(\begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{I} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 & 0 \\ 0 & \mathbf{B}_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} \\
&= \begin{bmatrix} p_{A_{k_1=k}} & p_{B_{k_1=k}} \\ p_{A_{k_2=k}} & p_{B_{k_2=k}} \end{bmatrix}
\end{aligned} \quad (1.55)$$

Note that the product \mathbf{BD} of the array **B** with the matrix **D** is interpreted as the $\kappa \times \kappa$ array with $d_{k'k} B_{k'}$ as entry in row k' and column k . By iterating (1.55), one arrives at the formula for arbitrary L :

$$\mathbf{p}_{AB} = \mathbf{Z}' \left((\mathbf{I} + \mathbf{BD}^{(1)}) \dots (\mathbf{I} + \mathbf{BD}^{(L-1)}) \right) \mathbf{V}. \quad (1.56)$$

The mean choice response time can be written accordingly for $L = 2$ as

$$\mathbf{et}_{AB} = \mathbf{Z}'((\mathbf{CD})\mathbf{V} + (\mathbf{I} + \mathbf{BD})\mathbf{W}). \quad (1.57)$$

and for arbitrary L as

$$\mathbf{et}_{AB} = \mathbf{Z}' \left[\left((\mathbf{CD}^{(1)}) \dots (\mathbf{CD}^{(L-1)}) \right) \mathbf{V} + ((\mathbf{I} + \mathbf{BD}) \dots (\mathbf{I} + \mathbf{BD})) \mathbf{W} \right]. \quad (1.58)$$

Combining (1.56) with (1.58) yields a joint recursion for computing \mathbf{p}_{AB} and \mathbf{et}_{AB} :

$$\begin{aligned} [\mathbf{p}_{AB}, \mathbf{et}_{AB}] &= [\mathbf{Z}', \mathbf{Z}'] \begin{bmatrix} \mathbf{I} + \mathbf{BD}^{(1)} & \mathbf{0} \\ \mathbf{CD}^{(1)} & \mathbf{I} + \mathbf{BD}^{(1)} \end{bmatrix} \\ &\quad \cdots \begin{bmatrix} \mathbf{I} + \mathbf{BD}^{(L-1)} & \mathbf{0} \\ \mathbf{CD}^{(L-1)} & \mathbf{I} + \mathbf{BD}^{(L-1)} \end{bmatrix} \begin{bmatrix} \mathbf{V} \\ \mathbf{W} \end{bmatrix}. \end{aligned} \quad (1.59)$$

In previous sequential decision models with finite L , the last attribute was always considered infinitely long (infinite decision horizon) to avoid the situation of no decision, i.e., $p_0 > 0$. This can be incorporated into the current model by modifying the definition of the matrices \mathbf{V}_k , \mathbf{W}_k corresponding to the last interval $[t_{L-1}, \infty)$ to

$$\mathbf{V}_k = (\mathbf{I} - \mathbf{Q}_k)^{-1} \mathbf{R}_k, \quad \mathbf{W}_k = (\mathbf{I} - \mathbf{Q}_k)^{-2} \mathbf{R}_k, \quad k = 1, \dots, K,$$

and modifying the recursion (1.59) slightly. Alternatively, one can artificially change the parameters of the attention time distribution for $l = L$ such that its expected value is sufficiently large, and make p_0 practically negligible.

References

- Borodin, A. N., & Salminen, P. (2002). *Handbook of Brownian motion – facts and formulae*, second edition. Basel: Birkhäuser Verlag.
- Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, 112, 117–128. doi:10.1037/0033-295X.112.1.117.
- Busemeyer, J., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences*, 43, 345–370.
- Busemeyer, J., & Townsend, J. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, 23(3), 255–282.
- Busemeyer, J., & Townsend, J. (1993). Decision field theory: A dynamic-cognitive approach to decision-making in an uncertain environment. *Psychological Review*, 100(3), 432–459.
- Carpenter, R. H. S., & Williams, M. L. L. (1995). Neural computation of log likelihood in the control of saccadic eye movements. *Nature*, 377, 59–62.
- Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, 39(2), 197–215.

- Diederich, A. (1997). Dynamic stochastic models for decision-making with time constraints. *Journal of Mathematical Psychology*, 41(3), 260–274.
- Diederich, A. (2003). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin and Review*, 10(1), 167–176.
- Diederich, A., & Oswald, P. (2014). Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Frontiers in Human Neuroscience*, 8(697).
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Karlin, S., & Taylor, H. (1981). *A second course in stochastic processes*. Orlando, FL: Academic Press.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Smith, P. (1995). Psychophysical principled models of visual simple reaction time. *Psychological Review*, 102, 567–591.
- Tuckwell, H. C. (1995). *Elementary applications of probability theory*, second edition. London: Chapman & Hall.
- Usher, M., & McClelland, J. (2001). The time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, 108, 550–592.
- Usher, M., & McClelland, J. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, 111, 757–769.

2 The Diffusion Model of Speeded Choice, from a Rational Perspective

Matt Jones

2.1	Binary Decision Tasks	71
2.2	Signal Detection Model	72
2.3	Random Walk Model	75
2.4	Continuous-time Model	80
2.5	Bayesian Diffusion Model	83
2.6	Translation between Diffusion Models	87
2.7	Predictions	92
2.8	Intertrial Variability and Unfalsifiability	97
2.9	Further Reading	100
2.10	Conclusions	101
	References	102

2.1 Binary Decision Tasks

This chapter considers tasks in which an experimental subject must make a binary decision, such as a perceptual discrimination or a semantic or mnemonic judgment. For some examples, a character could be presented on a monitor and the subject must decide whether it is red or green, or a letter or a number; or a string of letters is presented and the subject must decide whether it is a word or not; or a word is presented and the subject must decide whether it was part of a previously studied list or not. We are interested in both the probability that the subject will give the correct answer and the response time (RT) of whatever answer the subject gives.

Our aim is to develop formal, mathematical models of this type of task that make predictions for response probability and RT. This chapter will begin by following the historical progression of models developed in the literature, from *signal detection* models to *random walk* models to *diffusion* models, before presenting some new variations and results regarding the last of these.

All of the models considered here are based on the idea of *evidence sampling*. The assumption is that the subject, in perceiving the stimulus, observes or calculates some sort of information that bears on the correct answer. In a color discrimination task, this information would presumably concern the wavelength of

the light coming from the stimulus, originating in the subject's photoreceptors and further processed in the visual cortex, for example in red–green opponent-process cells. In a recognition memory task, the information would come from comparing the stimulus to memory, perhaps retrieving an explicit memory of the item from the study phase, or perhaps generating a continuous-valued familiarity signal. For the present purposes, we will not be concerned with the specific nature of this information or how it is computed. Instead, the focus will be on how the observations, once obtained, are used to generate a response in the binary decision task.

Under this view, the problem facing the subject is to determine the relative support that the observed information lends to the two responses. The models considered here take a normative approach to this problem, treating it as one of statistical inference. Under this approach, each stimulus category (i.e., correct response) is associated with a hypothesis. For example, hypothesis H_1 could be that the stimulus is red (Category 1), and hypothesis H_2 could be that the stimulus is green (Category 2). The subject's goal is to use the observations to infer which hypothesis is probably correct and to select the corresponding response. More specifically, the models here take a Bayesian approach, using the likelihood of the observations under each hypothesis to determine a posterior belief in which hypothesis is correct, which in turn drives decision-making.

2.2 Signal Detection Model

Consider first the simplest version of the model framework outlined above, where the subject observes a single sample and uses it to make a decision. This is referred to as a *signal detection model*, because historically it was developed for psychophysical tasks in which the subject's goal was to detect when a signal (e.g., an auditory tone) was presented, versus just background noise.

Define \mathcal{X} as the space of all possible observation values that could be observed, and $x \in \mathcal{X}$ as the value the subject actually observes. For each hypothesis, there exists a probability distribution over the value of x when that hypothesis is true (i.e., when a stimulus from that category is presented). For example, in a recognition memory task, if the observation takes the form of some familiarity signal, then we can define the distribution of familiarity values across all trials on which the stimulus is new (i.e., not on the studied list), as well as the distribution of familiarity values across all trials on which the stimulus was old. Formally, we write these two distributions as

$$P_1(x) = \Pr[x|H_1] \quad (2.1a)$$

and

$$P_2(x) = \Pr[x|H_2]. \quad (2.1b)$$

The notation $\Pr[x|H_i]$ indicates conditional probability, meaning the probability that x will be observed given that H_i is true.

Note that we need make no assumptions about the structure of the space \mathcal{X} . It could be a one-dimensional continuum (a subset of the real line), as in the case

of a recognition familiarity signal or net activation of red–green opponent cells. In richer perceptual tasks the space of observations could be multidimensional (a subset of \mathbb{R}^n), and in higher cognitive tasks like lexical decision, it could be some complex structured space of orthographic and semantic representations. The models considered here require only the functions P_1 and P_2 .¹

Considered as a function of the hypotheses, $P_i(x)$ (i.e., $P_1(x)$ vs. $P_2(x)$) is referred to as *likelihood*, and it determines the relative support that the observation lends to the hypotheses. The intuition is that, if x is more probable under Category 1 than Category 2, then observing x should increase the subject's belief that H_1 is the correct hypothesis. Assuming that the subject knows the functions $P_i(x)$ (or that these are the functions the subject believes, regardless of whether they are objectively accurate), then (s)he can use Bayes' rule to calculate the relative probabilities of the two hypotheses given the observation:

$$\begin{aligned}\frac{\Pr[H_1|x]}{\Pr[H_2|x]} &= \frac{\Pr[H_1]}{\Pr[H_2]} \cdot \frac{\Pr[x|H_1]}{\Pr[x|H_2]} \\ &= \frac{\Pr[H_1]}{\Pr[H_2]} \cdot \frac{P_1(x)}{P_2(x)}.\end{aligned}\quad (2.2)$$

The expression $\Pr[H_i|x]$ is called the *posterior probability* for H_i given x . This is the probability that the subject should assign to Category i after observing x (assuming inference is done optimally). The relation in (2.2) shows that the posterior probabilities of the two hypotheses depend on two things: the likelihoods and the *prior probabilities* $\Pr[H_i]$. The prior probabilities reflect the subject's beliefs about which response will be correct prior to observing the stimulus (i.e., before the start of the trial), for example, due to learning of base rates or of sequential patterns in the trial sequence.

If the subject's goal is to maximize the probability of choosing the correct answer, then the optimal decision rule is to select response R_1 if $\Pr[H_1|x] > \Pr[H_2|x]$, that is, if $\Pr[H_1|x]/\Pr[H_2|x] > 1$, and to select response R_2 otherwise (the choice is arbitrary in case of equality). In the simplest case where the priors are equal, $\Pr[H_1] = \Pr[H_2] = \frac{1}{2}$, this decision rule reduces to comparing $P_1(x)$ and $P_2(x)$ and choosing whichever hypothesis has the greater likelihood. One can imagine dividing the space \mathcal{X} into two regions, according to whether $P_1(x) > P_2(x)$ or vice versa (again, cases of equality are assigned arbitrarily) and associating each region to the corresponding response. This partitioning might be done in advance, so that the decision-making process reduces to determining which region the observation lies in and selecting the associated response.

Figure 2.1 gives an illustration of this model for a simple case where \mathcal{X} is a unidimensional continuum and P_1 and P_2 are both Gaussian distributions with equal variance. This *equal-variance signal detection model* is the simplest in the family

¹ The definitions in (2.1) are written assuming \mathcal{X} is a discrete space. \mathcal{X} could also be taken as a continuous space, with P_1 and P_2 representing probability density (with respect to some measure on \mathcal{X}) rather than probability mass. Everything that follows applies equally well to the continuous case as to the discrete case.

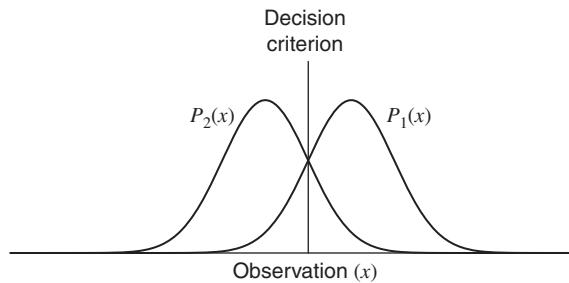


Figure 2.1 Illustration of the unbiased equal-variance signal detection model. An observation (x) is sampled from a unidimensional continuum represented by the horizontal line at the bottom. The probability distributions for x under the two hypotheses (i.e., stimulus categories) are represented by the curves labeled $P_1(x)$ and $P_2(x)$. Optimal decision-making amounts to determining where the observation lies relative to the decision criterion, which is the point where the likelihoods $P_1(x)$ and $P_2(x)$ are equal.

of models to be described in this chapter, and probably the most frequently applied in the psychological literature. In this model, there is a decision criterion that lies midway between the two distributions, at the point where $P_1(x) = P_2(x)$. Optimal decision-making in this model corresponds to selecting a response according to which side of the criterion the observation lies on.

We can generalize this simple model in three ways. First, we can allow arbitrary likelihood functions $P_i(x)$ on an arbitrary space of observations \mathcal{X} . Second, we can allow arbitrary values for the prior probability $\Pr[H_1]$ (with $\Pr[H_2] = 1 - \Pr[H_1]$). Third, we can introduce asymmetric reward structures, such that the reward for being correct or the penalty for being wrong is different for the two responses, by writing r_{ij} as the payoff for selecting response R_j when the correct response is R_i (with $r_{11} > r_{12}$ and $r_{22} > r_{21}$). Under this notation, the expected payoff for selecting R_j , conditioned on the observation, is given by

$$E[r|x, R_j] = r_{1j} \Pr[H_1|x] + r_{2j} \Pr[H_2|x]$$

and therefore the criterion for selecting response R_1 is:

$$E[r|x, R_1] > E[r|x, R_2] \iff \frac{\Pr[H_1|x]}{\Pr[H_2|x]} > \frac{r_{22} - r_{21}}{r_{11} - r_{12}}. \quad (2.3)$$

Combining this decision rule with (2.2) implies that R_1 is optimal if and only if

$$\frac{r_{11} - r_{12}}{r_{22} - r_{21}} \cdot \frac{\Pr[H_1]}{\Pr[H_2]} \cdot \frac{P_1(x)}{P_2(x)} > 1. \quad (2.4)$$

That is, the optimal response is determined by the net contribution of the prior belief, likelihood ratio, and ratio of reward contingency (i.e., the dependence of the reward on the subject's response, under each category). This is the optimal decision rule for the general signal detection model. In the unbiased case (equal priors and symmetric rewards), the ratios for the prior and reward contingency in

(2.4) equal unity, and we recover the simpler decision rule of selecting R_1 whenever $P_1(x) > P_2(x)$.

Finally, because of the multiplicative nature of the optimal decision rule, and because it depends on the observation only through the likelihood ratio, $P_1(x)/P_2(x)$, it is mathematically convenient to introduce the *log-likelihood ratio*:

$$L(x) = \ln \frac{P_1(x)}{P_2(x)}. \quad (2.5)$$

The log-likelihood ratio concisely captures the net evidence that the observation contributes to the two hypotheses. Taking the logarithm of (2.2) shows that $L(x)$ determines how much the subject's beliefs should change from prior to posterior, when those beliefs are expressed on a log-odds scale:

$$\ln \frac{\Pr[H_1|x]}{\Pr[H_2|x]} = \ln \frac{\Pr[H_1]}{\Pr[H_2]} + L(x). \quad (2.6)$$

Likewise, the criterion (2.4) for selecting R_1 can be re-expressed in logarithmic form,

$$\ln \frac{r_{11} - r_{12}}{r_{22} - r_{21}} + \ln \frac{\Pr[H_1]}{\Pr[H_2]} + L(x) > 0, \quad (2.7)$$

showing that the log reward contingency, prior log-odds, and log-likelihood ratio combine additively to determine the optimal response.

2.3 Random Walk Model

We now extend the signal detection model to assume that the subject observes not just one sample but a series of samples. For example, a subject in a perceptual discrimination task might process the stimulus multiple times in succession, or a subject in a recognition memory task might make multiple queries to memory. Formally, we model this observation process by assuming a series of observations x_n for $n = 1, 2, 3, \dots$, all jointly independent conditioned on the stimulus category and all sampled from the same distribution, P_1 or P_2 .

Denoting the sequence after n observations as $\mathbf{x}_n = (x_1, \dots, x_n)$, we can use the conditional independence assumption to write the likelihood as

$$\begin{aligned} \Pr[\mathbf{x}_n|H_i] &= \prod_{m=1}^n \Pr[x_m|H_i] \\ &= \prod_{m=1}^n P_i(x_m). \end{aligned} \quad (2.8)$$

Therefore, the log-likelihood ratio between the two hypotheses is equal to

$$\begin{aligned} \ln \frac{\Pr[\mathbf{x}_n|H_1]}{\Pr[\mathbf{x}_n|H_2]} &= \sum_{m=1}^n \ln \frac{P_1(x_m)}{P_2(x_m)} \\ &= \sum_{m=1}^n L(x_m). \end{aligned} \quad (2.9)$$

That is, the evidence provided by the sequence of observations is equal to the sum of the evidence provided by all of the individual observations. This is another important mathematical convenience of the log-odds representation, in addition to the additive relations in (2.6) and (2.7). Henceforth we use the term *evidence* specifically to refer to information or beliefs quantified on the log-odds scale.

Paralleling the derivation for the signal detection model, Bayes' rule gives the posterior odds, conditioned on the first n observations, as

$$\begin{aligned}\frac{\Pr[H_1|\mathbf{x}_n]}{\Pr[H_2|\mathbf{x}_n]} &= \frac{\Pr[H_1]}{\Pr[H_2]} \cdot \frac{\Pr[\mathbf{x}_n|H_1]}{\Pr[\mathbf{x}_n|H_2]} \\ &= \frac{\Pr[H_1]}{\Pr[H_2]} \cdot \prod_{m=1}^n \frac{P_1(x_m)}{P_2(x_m)}.\end{aligned}\quad (2.10)$$

As above, this relationship can also be written in terms of log-odds:

$$\ln \frac{\Pr[H_1|\mathbf{x}_n]}{\Pr[H_2|\mathbf{x}_n]} = \ln \frac{\Pr[H_1]}{\Pr[H_2]} + \sum_{m=1}^n L(x_m).\quad (2.11)$$

That is, the posterior log-odds equals the prior log-odds plus the sum of the log-likelihood ratios of the observations.

The expression in (2.11) suggests an intuitive process-level psychological model of decision-making. First, define the net evidence E_n as the posterior log-odds after n observations:

$$E_n = \ln \frac{\Pr[H_1|\mathbf{x}_n]}{\Pr[H_2|\mathbf{x}_n]}.\quad (2.12)$$

Thus, E_n represents the strength of belief that an optimal decision-maker will have in H_1 versus H_2 after observing x_1 through x_n . The expression for the posterior log-odds in (2.11) then implies a recursive relationship for E ,

$$E_n = E_{n-1} + L(x_n),\quad (2.13)$$

for all $n \geq 1$. Under the convention $\mathbf{x}_0 = \emptyset$ (i.e., an empty set of observations after zero trials), the starting point for E is the prior log-odds:

$$E_0 = \ln \frac{\Pr[H_1]}{\Pr[H_2]}.\quad (2.14)$$

Thus, the decision-maker starts at an evidence level determined by prior beliefs (E_0), and then uses the evidence provided by each successive observation ($L(x_n)$) to increment the cumulative evidence total (E_n).

Because x_n is a random variable (sampled from P_i when the stimulus comes from Category i), so is $L(x_n)$. This makes E a *stochastic process*, meaning a sequence of jointly distributed random variables. It is in fact a *Markov process*, meaning that the probability distribution for the next member of the sequence is fully determined by the value of the most recent member:

$$\Pr[E_n | \{E_0, \dots, E_{n-1}\}] = \Pr[E_n | E_{n-1}].\quad (2.15)$$

This relationship follows immediately from (2.13) together with the conditional independence between x_n and x_m for all $m < n$. Intuitively, a Markov process is thought of as memoryless, because its history has no impact on its future once its present state is known. This property makes this model appealing as a psychological model, because it implies that an optimal decision-maker needs only to track E_n from one observation to the next, rather than remembering the full sequence of past observations \mathbf{x}_n .

Geometrically, E can be conceived as a unidimensional random walk, wherein the posterior log-odds starts at the prior log-odds and moves up or down according to the evidence (i.e., log-likelihood ratio) provided by each successive observation. A (stationary) random walk is a Markov process where the probability distribution for the increment $E_n - E_{n-1}$ is independent of E_{n-1} and is the same for all n . In this case, the distribution for the increment $L(x_n)$ is independent of n because of the assumption that all observations are drawn from the same distribution (P_1 or P_2). It is also useful to consider the expected value of this increment, which determines how rapidly the evidence grows, on average. This is the *drift rate* of the random walk. When the stimulus is from Category 1, the drift rate is equal to

$$\begin{aligned} \mathbb{E}[L(x)|H_1] &= \mathbb{E}_{P_1}\left[\ln \frac{P_1(x)}{P_2(x)}\right] \\ &= D_{\text{KL}}(P_1\|P_2), \end{aligned} \quad (2.16a)$$

where \mathbb{E}_{P_1} indicates expected value according to the distribution P_1 , and D_{KL} denotes *Kullback–Leibler (KL) divergence*, a standard measure of the difference between two probability distributions ($D_{\text{KL}}(p\|q) = \mathbb{E}_p\left[\ln \frac{p}{q}\right]$). Likewise, the drift rate when the stimulus is from Category 2 is given by

$$\begin{aligned} \mathbb{E}[L(x)|H_2] &= \mathbb{E}_{P_2}\left[\ln \frac{P_1(x)}{P_2(x)}\right] \\ &= -D_{\text{KL}}(P_2\|P_1). \end{aligned} \quad (2.16b)$$

These are sensible results, because they mean that the more different the two hypotheses are (i.e., the greater the divergence between P_1 and P_2 or vice versa), the faster an optimal observer will typically be able to discriminate between them based on a sequence of samples from one or the other.

To take an example, consider the Gaussian equal-variance signal detection model (Figure 2.1), generalized from one observation to a sequence of observations. Let μ_1 and μ_2 be the respective means of the Gaussian distributions P_1 and P_2 , and let σ^2 be their shared variance. For a given observation x , the log-likelihood ratio is equal to

$$\begin{aligned} L(x) &= \ln \frac{\exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)} \\ &= \frac{\mu_1 - \mu_2}{\sigma^2} \left(x - \frac{\mu_1 + \mu_2}{2}\right). \end{aligned} \quad (2.17)$$

Thus, the evidence provided by an observation is a linear function of the value of that observation, with slope proportional to the mean difference between the two distributions and with neutral point (i.e., $L(x) = 0$) at the midpoint between the distributions. Because $L(x)$ is a linear function of a Gaussian variable, it has a Gaussian distribution as well. From (2.17), the mean of $L(x)$ is equal to

$$\begin{aligned} \mathbb{E}[L(x)|H_i] &= \frac{\mu_1 - \mu_2}{\sigma^2} \left(\mathbb{E}[x|H_i] - \frac{\mu_1 + \mu_2}{2} \right) \\ &= \pm \frac{1}{2} \left(\frac{\mu_1 - \mu_2}{\sigma} \right)^2 \end{aligned} \quad (2.18)$$

(with positive sign for $i = 1$ and negative sign for $i = 2$), and its variance is equal to

$$\begin{aligned} \text{var}[L(x)|H_i] &= \left(\frac{\mu_1 - \mu_2}{\sigma^2} \right)^2 \text{var}[x|H_i] \\ &= \left(\frac{\mu_1 - \mu_2}{\sigma} \right)^2. \end{aligned} \quad (2.19)$$

The ratio $(\mu_1 - \mu_2)/\sigma$ is referred to as d' in signal detection theory and is used as a measure of the standardized difference between distributions under the two hypotheses (most frequently when those distributions are assumed to be Gaussian with equal variance). The derivation in (2.18) shows that the drift rate of the random walk is proportional to the square of this measure, consistent with the intuition that the mean rate of evidence accumulation is greater with greater separation between the distributions.

The absolute value of the drift rate is equal under the two hypotheses in the equal-variance Gaussian model, but it should be noted that this is not true in general. This is related to the fact that KL divergence is generally asymmetric. For example, in a model with Gaussian sampling distributions of unequal variances (σ_1^2 and σ_2^2), the drift rates turn out to be

$$D_{\text{KL}}(P_1 \| P_2) = \frac{1}{2} \left(\left(\frac{\mu_1 - \mu_2}{\sigma_2} \right)^2 - 1 + \frac{\sigma_1^2}{\sigma_2^2} - \ln \frac{\sigma_1^2}{\sigma_2^2} \right) \quad (2.20a)$$

and

$$-D_{\text{KL}}(P_2 \| P_1) = -\frac{1}{2} \left(\left(\frac{\mu_1 - \mu_2}{\sigma_1} \right)^2 - 1 + \frac{\sigma_2^2}{\sigma_1^2} - \ln \frac{\sigma_2^2}{\sigma_1^2} \right), \quad (2.20b)$$

which are generally not equal in absolute value. If, say, $\sigma_1 > \sigma_2$, then the drift rate under Category 1 will be greater in magnitude than the drift rate under Category 2. Intuitively, this is because it is easier for extreme samples from Category 1 to provide strong evidence against H_2 than vice versa.

The random walk model can be applied as a psychological model of decision-making under two additional assumptions. First, there must be some decision rule that specifies at every possible state of the process – that is, for any value of the pair

(n, E_n) – whether the subject terminates the sampling process and responds with R_1 or R_2 , or whether the subject defers the decision and continues with another sample. Second, a time constant Δt must be specified, representing the physical time elapsed between successive samples (for simplicity, we assume samples are equally spaced). Under specifications of these assumptions, the model's prediction on any trial comprises the response dictated by the decision rule together with the response time (RT). The predicted RT will equal $n \cdot \Delta t$, where n is the number of samples observed, plus perhaps some non-decision time t_0 to model processes such as sensory encoding and motor execution that occur outside of the decision process itself. Across trials, the response and RT are jointly distributed random variables. Thus, the model's predictions constitute a joint distribution over the response and RT, separately for each stimulus category. Equivalently, for each stimulus category, the model yields predictions for the probabilities of both responses together with a conditional distribution of RT for each response (i.e., correct and incorrect).

An appealing feature of this psychological model is that it corresponds to the classic sequential probability-ratio test (SPRT), a statistical procedure wherein a sequence of observations is used to decide between two hypotheses. A central question in that statistical setting concerns the optimal stopping rule: when should the observer stop sampling and make a decision, versus continuing to draw more samples? A standard result known as the Wald–Wolfowitz theorem states that the optimal decision rule for the SPRT is to sample until the posterior reaches either an upper threshold, α , or a lower threshold, β (typically $\beta < 0 < \alpha$), and then to choose R_1 or R_2 accordingly. Both thresholds are fixed across time, that is, independent of n . Figure 2.2 illustrates the random walk model with this decision rule. According to the Wald–Wolfowitz theorem, this model is optimal in the sense that any other decision rule with the same error rates (i.e., probabilities of choosing R_2 when H_1 is true and of choosing R_1 when H_2 is true) will require at least as many samples on average. Thus, it is impossible to achieve superior accuracy with a shorter mean RT. In particular, one can set the thresholds as α and $-\alpha$ in cases where it is desirable to equate the two error rates, for example because they incur equal costs (i.e., the reward contingency ratio in (2.3) equals unity). The choice of α determines the observer's *speed–accuracy trade-off*, in that smaller α leads to faster responses but larger α leads to fewer errors. We defer analysis of RT predictions until Section 2.7, but it is easy to see that the log-odds of an error will approximately equal $-\alpha$, or equivalently the error rates will both be approximately $(1 + e^\alpha)^{-1}$. This is because the observer is terminating the trial when the posterior probability of the chosen response being correct is approximately α (assuming the observer has access to the correct prior probability and likelihood functions, and so can compute the correct posterior). This value for the error rate is nonetheless approximate because, with a discrete sequence of samples, the random walk will generally jump across the threshold rather than landing exactly on it. In the continuous-time model that we consider next, the approximation becomes exact.

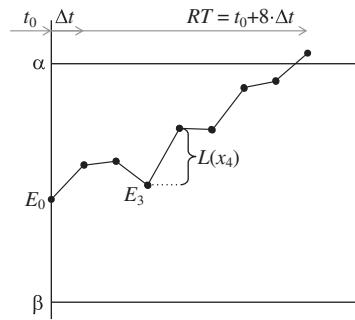


Figure 2.2 Illustration of the random walk model. Evidence (E) for deciding between two hypotheses is accumulated over time based on a series of observations (x_n). The evidence contributed by each observation is equal to its log-likelihood ratio ($L(x_n)$). The starting point for the evidence process (E_0) is determined by the prior probabilities of the hypotheses, specifically as the prior log-odds. After n observations, the posterior log-odds is given by E_n . Each hypothesis has a threshold, α for H_1 and β for H_2 , and crossing this threshold will trigger the sampling process to stop and the corresponding response to be selected. The temporal spacing between successive observations is denoted Δt , and the model's response time (RT) is equal to that spacing times the total number of observations taken, plus a non-decision time (t_0).

2.4 Continuous-time Model

We now build on the random walk model to derive a model in which evidence accumulation occurs not in discrete steps but continuously. To achieve this, we consider a sequence of random walk models in which the time between samples (Δt) approaches zero, and we derive a continuous-time evidence process that is the limit of the discrete-time random walks. Under the right assumptions about the means and variances of the evidence increments in the discrete models, the limiting process follows a directed Wiener diffusion (i.e., Brownian motion) process. This limiting model, when viewed as a psychological model of decision-making, is known as the *diffusion model*.

Formally, assume there exists a sequence of random walk models, indexed by $k \in \mathbb{N}$, with $\lim_{k \rightarrow \infty} \Delta t_k = 0$ and with sampling distributions P_i^k ($i \in \{1, 2\}$) and corresponding log-likelihood functions $L_k(x)$ that satisfy

$$\mathbb{E}[L_k(x)|H_i] = \xi_i \Delta t_k \quad (2.21)$$

and

$$\text{var}[L_k(x)|H_i] = \eta^2 \Delta t_k. \quad (2.22)$$

That is, both the conditional means and the conditional variance of the evidence increments in each model are proportional to its time step. Thus, the mean rate of evidence accumulation, $\mathbb{E}[L(x)|H_i]/\Delta t$, and the growth rate of the variance,

$\text{var}[L(x)|H_i]/\Delta t$, are both constant across all models in the sequence. This property is critical for the continuous-time limit to be well-behaved, as will be seen shortly.

It is easy to construct such a sequence of models. For example, a sequence of equal-variance Gaussian sampling models can be defined by setting

$$d'_k = c\sqrt{\Delta t_k} \quad (2.23)$$

for any constant $c > 0$. (Recall that $d' = (\mu_1 - \mu_2)/\sigma$; therefore (2.23) could be achieved in various ways, for example by fixing μ_1 and σ and varying μ_2 .) Thus, this construction implies that the discriminability between P_1 and P_2 converges to zero in proportion to the square-root of the time step. From (2.18) and (2.19), the mean and variance of the evidence increments are

$$\mathbb{E}[L_k(x)|H_i] = \pm \frac{c}{2}\Delta t_k \quad (2.24)$$

and

$$\text{var}[L_k(x)|H_i] = c\Delta t_k. \quad (2.25)$$

Therefore, we can take any sequence of Δt_k satisfying $\lim_{k \rightarrow \infty} \Delta t_k = 0$, and the random walk models defined by (2.23) will satisfy the scaling properties in (2.21) and (2.22).

Given any sequence of random walk models satisfying (2.21) and (2.22), all with the same prior log-odds, define each model's evidence trajectory through time as

$$E_k(t) = E_{\lfloor t/\Delta t_k \rfloor}^k, \quad (2.26)$$

where $\lfloor \cdot \rfloor$ is the floor function (i.e., $\lfloor z \rfloor$ is the greatest integer less than or equal to z), and E_n^k is the evidence level (i.e., posterior log-odds) of model k after n observations as in (2.12). The definition in (2.26) formalizes the assumption that the model's evidence is updated after each Δt_k interval, by setting $E_k(t)$ equal to the value of the evidence process after the last update before time t . Note that $E_k(0)$ is equal to the common prior log-odds, which we denote as $E(0)$. For each k , the trajectory $E_k(t)$ is a random function of t , with distribution governed by the updating rule in (2.13) and by the distribution of increments $L_k(x)$. The question now is how the distribution of $E_k(t)$ behaves under the limit $k \rightarrow \infty$.

For any single value of t , $E_k(t)$ is equal to $E(0)$ plus a sum of log-likelihood increments from $\lfloor t/\Delta t_k \rfloor$ independent observations, with the conditional mean and variance of these increments given by (2.21) and (2.22). Because the expressions in (2.21) and (2.22) are proportional to Δt_k , the mean and variance of the sum of the increments depends only on t and not on Δt_k , except for contributions of rounding error in $\lfloor t/\Delta t_k \rfloor$. Therefore, in the limit, the distribution of $E_k(t)$ obeys

$$\begin{aligned} \lim_{k \rightarrow \infty} \mathbb{E}[E_k(t)|H_i] &= \lim_{k \rightarrow \infty} \left(E(0) + \xi_i \Delta t_k \left\lfloor \frac{t}{\Delta t_k} \right\rfloor \right) \\ &= E(0) + \xi_i t \end{aligned} \quad (2.27)$$

and

$$\begin{aligned}\lim_{k \rightarrow \infty} \text{var}[E_k(t) | H_i] &= \lim_{k \rightarrow \infty} \left(\eta^2 \Delta t_k \left\lfloor \frac{t}{\Delta t_k} \right\rfloor \right) \\ &= \eta^2 t.\end{aligned}\quad (2.28)$$

Moreover, the central limit theorem implies that the limiting distribution is Gaussian. The same considerations apply to the difference $E_k(t_2) - E_k(t_1)$ for any two time points $t_1 < t_2$, and because of the Markov property of the random walk models these properties hold conditioned on the history of the process up to t_1 , denoted $\mathbf{E}_k(t_1)$:

$$\lim_{k \rightarrow \infty} \mathbb{E}[E_k(t_2) - E_k(t_1) | H_i, \mathbf{E}_k(t_1)] = \xi_i (t_2 - t_1) \quad (2.29)$$

and

$$\lim_{k \rightarrow \infty} \text{var}[E_k(t_2) - E_k(t_1) | H_i, \mathbf{E}_k(t_1)] = \eta^2 (t_2 - t_1), \quad (2.30)$$

with a distribution that is Gaussian in the limit.

The properties in (2.27)–(2.30) imply that the processes $E_k(t)$ (conditioned on each hypothesis H_i) converge in distribution to a Wiener diffusion process, $E(t)$, with drift rate ξ_i and diffusion rate η^2 . This Wiener diffusion process is defined as a stochastic process where the marginal distribution at any point in time is Gaussian with linear growth in the mean and variance,

$$E(t) \sim \mathcal{N}(E(0) + \xi_i t, \eta^2 t), \quad (2.31)$$

and with increments that are independent of the history (a property that can be seen to imply the Markov property):

$$E(t_2) - E(t_1) \perp \mathbf{E}(t_1) | H_i. \quad (2.32)$$

More precisely, for any finite set of time points $t_1 < \dots < t_m$, the values of the process $E(t_1)$ through $E(t_m)$ have a multivariate Gaussian distribution, with means and variances given by (2.31), and with covariances given by

$$\text{cov}[E(t_i), E(t_j)] = \eta^2 \min\{t_i, t_j\}. \quad (2.33)$$

Figure 2.3 illustrates the relationship between the discrete-time random walk model and the continuous-time diffusion model. Under the construction described here, one can think of the time step of the random walk model being repeatedly subdivided until (in the limit) the process evolves continuously. The linear scaling properties in (2.21) and (2.22) imply that this subdivision operation preserves the growth rate in both the expected value and the variance of the process, so that these rates are well-defined in the limit. The growth rate of the mean is referred to as the *drift rate*, and the growth rate of the variance as the *diffusion rate*.

The diffusion process can be built into a psychological model of speeded decision-making just like the random walk model was. The model embodies the idea that decisions are made based on a continuous stream of evidence, formally

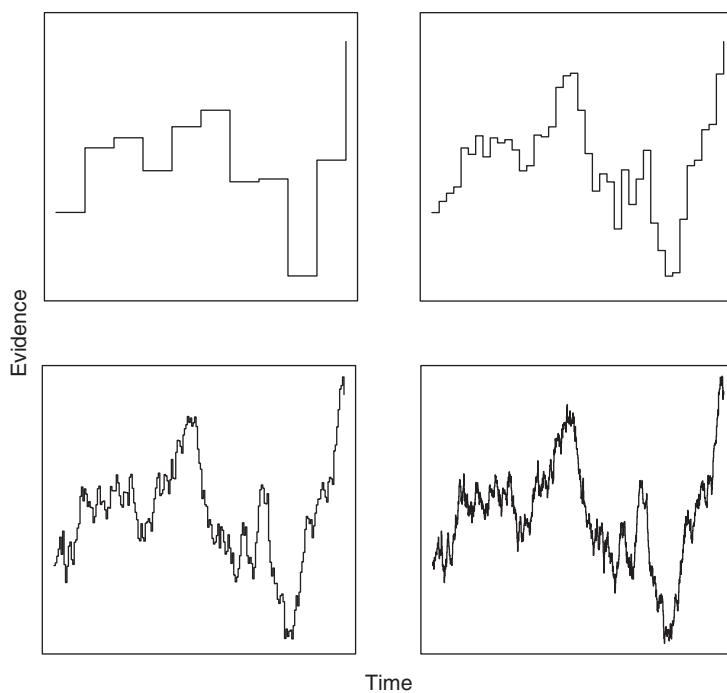


Figure 2.3 *Illustration of the relationship between random walk and diffusion models of evidence accumulation. Each plot shows a sample trajectory from a random walk process. Proceeding from upper left to lower right, the time step becomes finer and the evidence increments become smaller. If the mean and variance of the increments are proportional to the time step, then the sequence of random walks converges in distribution to a diffusion process.*

described by a Wiener process. Completing the model requires specifying upper and lower thresholds, α and β , for the two responses. Because the evidence process is continuous, these thresholds will exactly determine the log-odds of a correct response, unlike the random walk model where the evidence jumps across the threshold. That is, when the model selects response R_1 , the log-odds that Category 1 is correct exactly equals α , and similarly with R_2 and β (again, this assumes the observer knows the correct prior $E(0)$ and log-likelihood ratio function $L(x)$). Also, the diffusion model predicts a dense set of possible RT values, as opposed to the discrete set of possible RTs under a pure random walk model.

2.5 Bayesian Diffusion Model

Although the historic development of random walk and diffusion models of decision-making was founded on the framework of statistical inference in the SPRT, current treatments of the diffusion model in the psychological literature depart from the normative framing, casting the diffusion process at a purely

mechanistic level. That is, they simply posit that decision-making is based on some cognitive variable that obeys the dynamics of a diffusion process, without interpreting that process as the result of Bayesian inference over some input stream. Here we show how this mechanistic diffusion model can be reformulated in the normative framework of Section 2.4, and how such a reformulation offers further insights into the properties of the model.

The starting point of the mechanistic diffusion model is to assume a stochastic process, denoted here as $e(t)$ for $t \geq 0$, defined by the stochastic differential equation

$$de = \mu_i dt + \sigma dB(t). \quad (2.34)$$

Here, i indicates the true stimulus category, and $B(t)$ represents a standard Brownian motion process (i.e., with zero drift and with diffusion rate equal to unity). We assume without loss of generality that the input drift rates satisfy $\mu_1 > \mu_2$, and that $e(0) = 0$. The general version of the mechanistic diffusion model assumes the starting point $e(0)$ can take on arbitrary values, as a free parameter. This assumption is reintroduced in the next section, but it is irrelevant for the Bayesian derivation in the present section because the starting point can simply be subtracted away (e.g., one could define $e'(t) = e(t) - e(0)$ and do inference from e').

The Brownian motion process can be thought of as a limit of random walks with time step approaching zero (just as in Section 2.4). It has the properties that any increment $B(t_2) - B(t_1)$ is a random variable with normal distribution, zero mean, and variance equal to $t_2 - t_1$ (for $t_1 \leq t_2$); and that the value of such an increment is independent of the prior history, $(B(\tau))_{\tau \leq t_1}$. Thus, $e(t)$ evolves according to a sum of a deterministic, linear process with slope μ_i and a stochastic process with diffusion rate σ^2 . To be clear, we use the term *diffusion rate* to refer to σ^2 (not σ), because it represents the rate at which the variance in e grows over any interval of time:

$$\text{var}[e(t_2) - e(t_1)] = \sigma_i^2 (t_2 - t_1) \quad (2.35)$$

for $t_1 \leq t_2$. Typically, the diffusion rate is assumed to be the same under both stimulus categories, and below we give a justification for that assumption based on the Bayesian interpretation.

Our Bayesian approach is to treat e as comprising the observations that the subject uses to infer the correct hypothesis. This is consistent with the mechanistic diffusion model, except that whereas that model directly applies an arbitrary threshold to $e(t)$, here we use $e(t)$ to calculate posterior log-odds and define the decision rule on the log-odds. Thus, $e(t)$ is a continuous analogue of x_n from the random walk model. More precisely, we write $\mathbf{e}_t = (e(\tau))_{\tau \leq t}$ as the full trajectory up to time t , and do Bayesian inference using \mathbf{e}_t in the same way the random walk model does inference using the discrete sequence \mathbf{x}_n .

We calculate a posterior from \mathbf{e}_t as follows. For any finite value of Δt , define a discrete sequence of observations x_n as increments of e :

$$x_n = e(n\Delta t) - e((n-1)\Delta t). \quad (2.36)$$

The properties of Brownian motion given above imply that these observations are jointly independent for different n and distributed as

$$x_n \sim \mathcal{N}(\mu_i \Delta t, \sigma^2 \Delta t). \quad (2.37)$$

The log-likelihood ratio for x_n can be calculated as in (2.17), as

$$L(x_n) = \frac{\mu_1 - \mu_2}{\sigma^2} \left(x_n - \frac{\mu_1 + \mu_2}{2} \Delta t \right). \quad (2.38)$$

Therefore, the log-likelihood ratio for the sequence $\mathbf{x}_n = (x_1, \dots, x_n)$ equals

$$\begin{aligned} \ln \frac{p[\mathbf{x}_n|H_1]}{p[\mathbf{x}_n|H_2]} &= \sum_{m=1}^n \frac{\mu_1 - \mu_2}{\sigma^2} \left(x_m - \frac{\mu_1 + \mu_2}{2} \Delta t \right) \\ &= \frac{\mu_1 - \mu_2}{\sigma^2} \left(e(n\Delta t) - \frac{\mu_1 + \mu_2}{2} n\Delta t \right). \end{aligned} \quad (2.39)$$

(We use p to indicate probability density, rather than \Pr for probability mass, because x and hence \mathbf{x} are now necessarily continuous-valued.) If we fix t and let $\Delta t = t/n$, then this equation becomes

$$\ln \frac{p[\mathbf{x}_n|H_1]}{p[\mathbf{x}_n|H_2]} = \frac{\mu_1 - \mu_2}{\sigma^2} \left(e(t) - \frac{\mu_1 + \mu_2}{2} t \right). \quad (2.40)$$

Thus, we see that inference does not depend on the step size Δt , and by letting $\Delta t \rightarrow 0$ (i.e., $n \rightarrow \infty$), that $e(t)$ is a sufficient statistic for inferring the correct hypothesis from the full trajectory of the process, \mathbf{e}_t . That is, the posterior probability depends only on the current value of the diffusion process and not on its history.

Therefore, we have a well-defined continuous-time Bayesian evidence-accumulation model, with input defined by the stochastic process $e(t)$. Using the expression for the log-likelihood ratio in (2.40), and the invariance across sampling densities (i.e., Δt), the posterior log-odds are given by

$$\ln \frac{\Pr[H_1|\mathbf{e}_t]}{\Pr[H_2|\mathbf{e}_t]} = \ln \frac{\Pr[H_1]}{\Pr[H_2]} + \frac{\mu_1 - \mu_2}{\sigma^2} \left(e(t) - \frac{\mu_1 + \mu_2}{2} t \right). \quad (2.41)$$

Therefore, the posterior is a linear transformation of the input, and thus follows a diffusion process itself, with new drift and diffusion rates.

To understand the relationship between the mechanistic diffusion model (where decisions are based directly on the input $e(t)$) and the Bayesian model (where the input is first transformed according to (2.41)), we reparameterize the latter as follows. First, define a *drift criterion*,

$$\theta = \frac{\mu_1 + \mu_2}{2}, \quad (2.42)$$

as the midpoint between the input drift rates for the two categories. The drift criterion is a dynamic analogue of the equal-likelihood criterion in signal-detection

theory, in that $e(t) > \theta t$ implies a positive log-likelihood ratio (i.e., evidence for H_1), and $e(t) < \theta t$ implies a negative log-likelihood ratio (i.e., evidence for H_2). Second, define a *signal-to-noise ratio*,

$$\phi = \frac{\mu_1 - \mu_2}{\sigma^2}, \quad (2.43)$$

as the difference in input drift rates between hypotheses divided by the diffusion rate. The signal-to-noise ratio is analogous to the concept of d' in signal detection theory in that it gives a standardized measure of how separated are the sampling distributions under the two hypotheses. Another connection between d' and ϕ is that both parameters determine how informative the observations are, by providing scaling factors to convert from the input to log-likelihood ratio. From (2.17), the log-likelihood ratio in Gaussian equal-variance signal detection and random walk models is $L(x) = d' \cdot (x - \theta) / \sigma$, whereas in the Bayesian diffusion model the log-likelihood ratio is $L(e(t)) = \phi \cdot (e(t) - \theta t)$. Thus, larger values of ϕ imply that the posterior moves more rapidly with changes in e .

As above, let $E(t)$ denote the evidence level (i.e., posterior log-odds) at time t :

$$E(t) = \ln \frac{\Pr[H_1 | \mathbf{e}_t]}{\Pr[H_2 | \mathbf{e}_t]}. \quad (2.44)$$

Then, using the above definitions, (2.41) becomes

$$E(t) = E(0) + \phi (e(t) - \theta t). \quad (2.45)$$

Therefore, optimal inference requires knowledge of three parameters: the prior probability ($E(0)$), the drift criterion (θ), and the signal-to-noise ratio (ϕ). These can be considered as properties of the task environment or the psychological processes generating the input signal $e(t)$. Provided the observer knows these three values, then calculating the posterior log-odds for the hypotheses can be accomplished by the linear transformation in (2.45). The result is a new diffusion process, with drift rates

$$\begin{aligned} \xi_i &= \phi (\mu_i - \theta) \\ &= \pm \frac{(\mu_1 - \mu_2)^2}{2\sigma^2}, \end{aligned} \quad (2.46)$$

(positive for $i = 1$ and negative for $i = 2$) and diffusion rate

$$\begin{aligned} \eta^2 &= \phi^2 \sigma^2 \\ &= \frac{(\mu_1 - \mu_2)^2}{\sigma^2}. \end{aligned} \quad (2.47)$$

If the observer wants to achieve a given accuracy level $\Pr[\text{correct}] = \rho$ (equal for the two categories), then the optimal decision rule is to terminate sampling and choose a response as soon as $|E(t)| \geq \alpha$, using a threshold given by

$$\alpha = \ln \frac{\rho}{1 - \rho}. \quad (2.48)$$

As in the random walk model, the choice of threshold controls the subject's speed-accuracy trade-off, with larger threshold values yielding greater accuracy but slower RT.

2.6 Translation between Diffusion Models

Because the evidence $E(t)$ itself follows a diffusion process, in practice one might dispense with e and use E directly as the starting point for modeling. From a mechanistic standpoint, e adds nothing to the model's predictions, barring some physiological theory of the input signal that would allow it to be measured. Indeed, the Bayesian diffusion model can be treated mathematically as simply a special case of the standard, mechanistic model: it is defined by a stochastic evidence process, $E(t)$, with particular values for the drift rates and diffusion rate, as well as a starting point $E(0)$ and thresholds $\pm\alpha$. As seen in (2.46) and (2.47), the drift rate of $E(t)$ is necessarily equal to half of its diffusion rate, and otherwise the model's parameters (i.e., starting point, threshold, and drift/diffusion rates) can be independently specified.

Therefore, we can think of the Bayesian diffusion model as a reparameterization of the mechanistic one. Under this view, the value of the derivation presented in Section 2.5 is that it provides a normative Bayesian foundation for the diffusion model. To summarize, under the Bayesian diffusion model, the observer is assumed to have access to a continuous stream of input information conforming to a Wiener diffusion process with unknown drift rate, or equivalently an integral of a white-noise process with unknown mean. Optimal Bayesian inference is applied to this input to infer the correct hypothesis via its influence on the input drift rate. The evidence process $E(t)$ is obtained from the input by subtracting the drift criterion, which corresponds to neutral input that has equal likelihood under both hypotheses; scaling by the signal-to-noise ratio, which determines how much information the input carries, to transform to units of log-odds; and adding the prior log-odds to reflect prior expectations. The drift and diffusion rates of this evidence process are determined by the difference in input rates between the two hypotheses, together with the noise (diffusion rate) in the input. The starting point, $E(0)$, is determined by the prior probabilities of the two stimulus categories. The decision thresholds correspond to the observer's choice of the log-odds that each response will be correct. Provided the observer knows the input rates μ_1 and μ_2 (or equivalently, θ and ϕ) and the prior probabilities (e.g., the base rate), then the inference process described above can be carried out, and any desired performance level ρ can be achieved optimally – that is, while minimizing the mean RT. Thus, the Bayesian diffusion model also raises the question of how an experimental subject might come to know these task parameters, which suggests rich opportunities for integrating diffusion models of decision-making with mathematical models of learning.

This normative framing offers answers to two conceptual challenges within the diffusion framework. First, one might ask why the diffusion model assumes the

same diffusion rate under both hypotheses. After all, this is not the case under the signal detection or random walk models, where the variance of the input or of the evidence increment can differ between the hypotheses. Nevertheless, the equal-variance assumption is universal in applications of the diffusion model, and the Bayesian framing given here offers a justification. To see this, consider an alternative model in which the input process is defined by

$$de = \mu_i dt + \sigma_i dB(t), \quad (2.49)$$

with unequal diffusion rates $\sigma_1^2 \neq \sigma_2^2$. As above, define a sequence of discrete observations by (2.36). These observations now have a variance that depends on the correct hypothesis:

$$x_n \sim \mathcal{N}(\mu_i \Delta t, \sigma_i^2 \Delta t). \quad (2.50)$$

The log-likelihood ratio implied by this unequal-variance model is given by

$$L(x) = -\frac{(x - \mu_1 \Delta t)^2}{2\sigma_1^2 \Delta t} + \frac{(x - \mu_2 \Delta t)^2}{2\sigma_2^2 \Delta t} - \ln \frac{\sigma_1}{\sigma_2}, \quad (2.51)$$

and the expected value of this quantity under each stimulus category equals

$$\mathbb{E}[L(x) | H_1] = \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2} \Delta t + \frac{1}{2} \left(\frac{\sigma_1^2}{\sigma_2^2} - 1 - \ln \frac{\sigma_1^2}{\sigma_2^2} \right) \quad (2.52a)$$

and

$$\mathbb{E}[L(x) | H_2] = -\frac{(\mu_1 - \mu_2)^2}{2\sigma_1^2} \Delta t - \frac{1}{2} \left(\frac{\sigma_2^2}{\sigma_1^2} - 1 - \ln \frac{\sigma_2^2}{\sigma_1^2} \right), \quad (2.52b)$$

paralleling the result for the unequal-variance discrete models in (2.20). Now fix t and let $\Delta t = t/n$. The expected posterior log-odds after time t is given by

$$\begin{aligned} \mathbb{E}[E(t) | H_1] &= \ln \frac{\Pr[H_1]}{\Pr[H_2]} + \sum_{m=1}^n \mathbb{E}[L(x_m) | H_1] \\ &= \ln \frac{\Pr[H_1]}{\Pr[H_2]} + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2} t + \frac{n}{2} \left(\frac{\sigma_1^2}{\sigma_2^2} - 1 - \ln \frac{\sigma_1^2}{\sigma_2^2} \right) \end{aligned} \quad (2.53a)$$

and

$$\mathbb{E}[E(t) | H_2] = \ln \frac{\Pr[H_1]}{\Pr[H_2]} - \frac{(\mu_1 - \mu_2)^2}{2\sigma_1^2} t - \frac{n}{2} \left(\frac{\sigma_2^2}{\sigma_1^2} - 1 - \ln \frac{\sigma_2^2}{\sigma_1^2} \right). \quad (2.53b)$$

If the variances are equal then the last main summand in both (2.53a) and (2.53b) equals zero, and the posterior log-odds follows a diffusion process with drift rate $\pm (\mu_1 - \mu_2)^2 / (2\sigma^2)$ as already found above. However, if the variances differ then the last term will be strictly positive in (2.53a) and strictly negative in (2.53b). As

$\Delta t \rightarrow 0$, meaning $n \rightarrow \infty$, this term approaches infinity, implying the observer approaches perfect certainty of the correct category. This is a property of statistical inference with diffusion processes: observation on an arbitrarily fine timescale provides no advantage in inferring the drift rate (because the final value of the process is a sufficient statistic for that parameter), but it provides perfect information for inferring the diffusion rate. Therefore, any diffusion model of decision-making must assume equal diffusion rates for the two categories in order to be sensible from a Bayesian perspective. Otherwise, an ideal observer would be able to respond with perfect accuracy, using an arbitrarily short amount of time. This is also why we stated above that the observer needs only to know θ , ϕ , and the prior probabilities of the categories: σ does not need to be known because it can be exactly inferred from the input.

Second, the Bayesian diffusion model offers an elegant solution to a particular redundancy among the parameters in the mechanistic diffusion model. Specifically, the mechanistic diffusion model suffers a *scaling degeneracy* due to the fact that the scale of $e(t)$ has no impact on the model's predictions. If the drift rate, square-root of the diffusion rate (i.e., σ), starting point, and thresholds for the mechanistic model were all multiplied by any positive number, the only consequence would be a change in the internal scaling of $e(t)$, and the model's predictions for the joint distribution of response and RT would be unchanged. In practice, many modelers have adopted a convention to fix σ at an arbitrary value (usually 0.1), or in related models to fix the sum of the drift rates $\mu_1 + \mu_2$, so as to remove this degeneracy when estimating parameter values from data. In the Bayesian diffusion model, this indeterminacy never arises, because the evidence process has a uniquely determined scale, defined by log-odds (and log-likelihood ratio) of the two hypotheses. This unique scaling manifests in the constraint noted above for the model's parameters, that the diffusion rate is exactly twice the drift rate. Thus, the Bayesian diffusion model offers a more principled solution to the scaling degeneracy than the arbitrary solutions just mentioned, specifically by constraining the drift and diffusion rates to obey a 1:2 ratio. It is easy to see that, given any parameterization of the mechanistic model, it can always be rescaled to fit this form by multiplying all parameters by ϕ . This implies that the 1:2 constraint is not a predictive constraint at all; that is, it does not make the Bayesian model any more restricted than the mechanistic model. Moreover, the derivation of (2.45) shows that this rescaling always puts the evidence process into units of log-odds, and therefore has the benefit of making all model parameters more interpretable. In particular, under the parameterization entailed by the 1:2 constraint, the model's starting point can be directly interpreted as the observer's subjective prior log-odds, and the thresholds can be directly interpreted as values chosen by the observer for the log-odds that each response will be correct.

Despite the close correspondence between the Bayesian and mechanistic diffusion models, there are some important differences. First, the mechanistic model is often applied to experimental settings involving multiple stimulus subtypes within each category, such as perceptual stimuli of varying salience or lexical stimuli of varying corpus frequency. These cases are usually modeled by assuming a

different drift rate for each subtype. The Bayesian model can easily incorporate this assumption, by allowing different drift rates in the input process. It does, however, raise the question of how this additional complexity in the input relates to the hypotheses held by the observer. One possibility is that the input has multiple drift rates, μ_{ij} , indexed by both stimulus category (i) and subtype (j), but that the observer's hypotheses assume a single drift rate per category, $\hat{\mu}_i$, perhaps equal to the frequency-weighted mean of the true drift rates across all subtypes. Another possibility is that the hypotheses incorporate the subtyping of stimuli, such that H_i specifies a mixture distribution over diffusion processes with different drift rates, μ_{ij} for all possible j . The latter possibility would make the calculation of the posterior more complex than the result in (2.45). Either way, the 1:2 condition offered above for resolving the mechanistic model's scaling degeneracy is too simple to be applicable to experiment designs with multiple stimulus subtypes, but there should be extensions of this idea that would achieve similar results.

A second set of differences lies in constraints on the parameters of the two models. On one hand, the mechanistic model is more general than the Bayesian model in that it can assume arbitrary drift rates for the two stimulus categories, μ_1 and μ_2 , whereas in the Bayesian model the drift rates must satisfy $\xi_2 = -\xi_1$. However, if the mechanistic model assumes asymmetric drift rates ($\mu_2 \neq -\mu_1$) then its decision rule is non-optimal in the sense of the Wald–Wolfowitz theorem. That is, its thresholds (which are invariant as criteria on $e(t)$) become time-varying when translated to criteria on $E(t)$ (i.e., on posterior log-odds). This follows from (2.45) and the fact that $\theta \neq 0$ when the mechanistic model's drift rates are asymmetric. This issue is well known in the literature on the diffusion model, and in practice it is often assumed that the drift rates are symmetric ($\mu_2 = -\mu_1$), under the rationale that the subject knows the correct drift criterion and has already adjusted for it (i.e., by transforming $e(t)$ to $e(t) - \theta t$). On the other hand, the Bayesian model is more general than the mechanistic one because it has two forms of response bias, one in the prior ($E(0)$) and the other in the thresholds (α and β , perhaps reflecting asymmetries in the reward structure). These two types of bias are not separately interpretable in the mechanistic model, where they both manifest as differences in the distances from the starting point to the two thresholds. The mechanistic model appears to have three free parameters for the starting point and thresholds (thus giving it two degrees of freedom to capture response bias, in addition to the overall difference between the two thresholds for capturing speed–accuracy trade-off), but as with the scaling invariance discussed above, there is a translation invariance whereby adding a constant to $e(t)$ (and hence to the starting point and both thresholds) has no impact on model predictions. In other words, the evidence process in the mechanistic model is defined only on an interval scale, whereas in the Bayesian model it has a well-defined zero point.

Considering these differences in parameter constraints or redundancy in the two models, full equivalence between the Bayesian and mechanistic diffusion models is possible only under special cases of each. Specifically, we consider now the mechanistic model with symmetric drift rates ($\mu_2 = -\mu_1$), and the Bayesian model with

symmetric thresholds ($\beta = -\alpha$). Effectively, we remove the mechanistic model's ability to implement suboptimal decision rules, and we limit the Bayesian model to response bias in the prior (and not in reward structure). Under these restrictions, we can derive exact translations between the parameters of the two models. The mechanistic parameterization we use is the one that is most common in the literature, with drift rates $\pm\mu$, diffusion rate σ^2 , starting point z , and thresholds of 0 and a (with $a > 0$). These four parameters confer three degrees of freedom because of the scaling degeneracy. For the Bayesian parameterization, we denote the drift rates as $\pm\xi$, the diffusion rate 2ξ , the starting point $E(0)$, and the thresholds $\pm\alpha$.

The derivation of (2.45) assumed for simplicity a starting point of $e(0) = 0$ but, as noted above, in general calculation of the posterior requires subtracting $e(0)$ from $e(t)$. Also, the assumption of symmetric drift rates in the mechanistic model implies $\theta = 0$. Therefore (2.45) becomes

$$E(t) = E(0) + \frac{2\mu}{\sigma^2} (e(t) - z) \quad (2.54)$$

(where we have substituted the definition of ϕ from (2.43) to write everything in terms of the basic parameters of the mechanistic model). Thus, we see as in (2.46) and (2.47) that $E(t)$ is a diffusion process with drift rate (for Category 1) equal to

$$\xi = \frac{2\mu^2}{\sigma^2} \quad (2.55)$$

and diffusion rate equal to twice this quantity. From (2.54), requiring the thresholds of the two models to agree implies

$$\alpha = E(0) + \frac{2\mu}{\sigma^2} (a - z) \quad (2.56a)$$

and

$$-\alpha = E(0) + \frac{2\mu}{\sigma^2} (-z). \quad (2.56b)$$

Subtracting these two equations provides the translation between the models' threshold parameters,

$$\alpha = \frac{\mu a}{\sigma^2}, \quad (2.57)$$

and then substituting back into (2.56) gives the translation between starting points:

$$E(0) = \frac{2\mu}{\sigma^2} \left(z - \frac{a}{2} \right). \quad (2.58)$$

Finally, we can substitute back into (2.54) to obtain a translation between the two diffusion processes in terms of only the mechanistic model's parameters:

$$E(t) = \frac{2\mu}{\sigma^2} \left(e(t) - \frac{a}{2} \right). \quad (2.59)$$

This result is sensible, because $a/2$ is the midpoint between thresholds and thus represents a neutral point corresponding to $E(t) = 0$, and because we have already

Table 2.1 *Translation between mechanistic and Bayesian versions of the diffusion model.*

Property	Mechanistic model	Bayesian model	Translation
Diffusion process	$e(t)$	$E(t)$	$E(t) = \frac{2\mu}{\sigma^2} (e(t) - \frac{a}{2})$
Drift rate	$\pm\mu$	$\pm\xi$	$\xi = \frac{2\mu}{\sigma^2}$
Diffusion rate	σ^2	2ξ	$2\xi = \frac{4\mu^2}{\sigma^2}$
Starting point	z	$E(0)$	$E(0) = \frac{2\mu}{\sigma^2} (z - \frac{a}{2})$
Thresholds	$\{0, a\}$	$\pm\alpha$	$\alpha = \frac{\mu a}{\sigma^2}$

Note: Models are restricted to assume homogeneous stimuli within each category (no stimulus subtypes), to use symmetric drift rates (i.e., mechanistic model assumes the observer has adjusted for the drift criterion), and symmetric thresholds for the Bayesian model (i.e., response bias lies only in the starting point). The translation column gives the unique translation from the mechanistic parameterization to the Bayesian one. The reverse translation is unique only up to multiplicative scaling of the mechanistic model's parameters, because of the scaling degeneracy in that model.

seen that $\phi = 2\mu/\sigma^2$ is the scaling factor that converts the units of $e(t)$ into units of log-odds. Table 2.1 summarizes the correspondences between the two models. Under this translation, the mechanistic and Bayesian diffusion models make identical predictions and differ only in theoretical interpretation.

2.7 Predictions

We now consider the diffusion model's predictions for response probability and RT. We do this for both the mechanistic and the Bayesian parameterizations of the model (see Table 2.1). According to the operation of the model, on each decision trial a diffusion process begins at the specified starting point and evolves stochastically until it reaches one threshold or the other. The model's response is determined by which threshold is crossed first, and its RT by the time it takes for that crossing to occur (plus perhaps some non-decision time). The model's prediction for the joint distribution of response and RT is determined by aggregating across the ensemble of possible trajectories for the diffusion process, as illustrated in Figure 2.4.

The model's response probabilities, $\Pr[R_j|H_i]$ for $i, j \in \{1, 2\}$, have been presented many times in the literature, but here we show that the Bayesian formulation provides an intuitive method for deriving these quantities. The key insight is that, because $E(t)$ represents the true posterior log-odds at any time t , the value of E at any time determines the objective probability that either response will be correct. Thus, if the model selects response R_1 because the process terminates at $E = \alpha$, then the log-odds that this response is correct equals α , and if the model selects R_2 because the process terminates at $E = -\alpha$, then the log-odds that that response

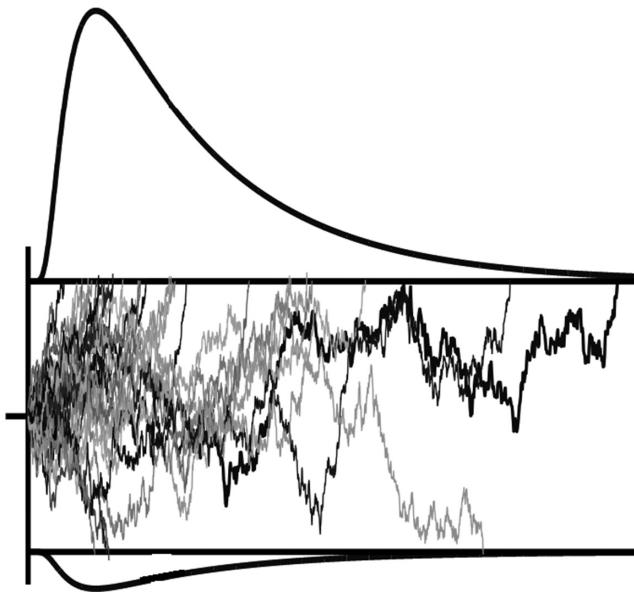


Figure 2.4 Illustration of the operation of the diffusion model. At the start of each decision trial (indicated by the vertical line, with time flowing to the right), a diffusion process begins at a value determined by the model's starting point parameter (horizontal tick). The process evolves stochastically until it reaches one of the two thresholds (horizontal lines). Greyscale curves show sample trajectories. The model's response and response time (RT) on each trial are determined by which threshold is crossed first and when the crossing occurs. The predicted joint distribution of response and RT is represented by the curves at top and bottom, which indicate the probability density of RT for each response. The area subsumed by each curve represents the marginal probability of that response.

is correct equals α and the log-odds that R_1 would have been correct equals $-\alpha$. Writing these dependencies in terms of probabilities, we have

$$\Pr [H_1|R_1] = \frac{1}{1 + e^{-\alpha}} \quad (2.60a)$$

and

$$\Pr [H_1|R_2] = \frac{1}{1 + e^{\alpha}}. \quad (2.60b)$$

From the definition of $E(0)$ as the prior log-odds, the marginal probabilities for the two hypotheses are given by

$$\Pr [H_1] = \frac{1}{1 + e^{-E(0)}} \quad (2.61a)$$

and

$$\Pr [H_2] = \frac{1}{1 + e^{E(0)}}. \quad (2.61b)$$

Combining the above four relations enables us to derive the conditional probability for the response given each hypothesis. We use the following basic identity for binary random variables, which can be derived from the definitions of joint and conditional probability:

$$\Pr [R_1] = \frac{\Pr [H_1] - \Pr [H_1|R_2]}{\Pr [H_1|R_1] - \Pr [H_1|R_2]}. \quad (2.62)$$

The conditional response probability under Category 1 can then be derived as

$$\begin{aligned} \Pr [R_1|H_1] &= \frac{\Pr [R_1] \Pr [H_1|R_1]}{\Pr [H_1]} \\ &= \frac{e^\alpha - e^{-E(0)}}{e^\alpha - e^{-\alpha}}, \end{aligned} \quad (2.63a)$$

and by analogous calculations, the response probability under Category 2 is

$$\Pr [R_1|H_2] = \frac{e^{E(0)} - e^{-\alpha}}{e^\alpha - e^{-\alpha}}. \quad (2.63b)$$

Notice that these results do not depend on the drift or diffusion rate of the Bayesian evidence process (ξ), which is a consequence of the 1:2 property of the Bayesian model, and which contrasts with the results for the mechanistic model given below. In the special case of a neutral prior ($E(0) = 0$, corresponding to $\Pr [H_1] = \Pr [H_2] = \frac{1}{2}$), the response probabilities reduce to

$$\Pr [R_1|H_1] = \frac{1}{1 + e^{-\alpha}} \quad (2.64a)$$

and

$$\Pr [R_1|H_2] = \frac{1}{1 + e^\alpha}, \quad (2.64b)$$

meaning that the log-odds of a correct response under either stimulus category equal α .

To derive the response probabilities for the mechanistic model, we substitute the parameter translations in Table 2.1 into (2.63), yielding

$$\Pr [R_1|H_1] = \frac{1 - e^{-\frac{2\mu z}{\sigma^2}}}{1 - e^{-\frac{2\mu a}{\sigma^2}}} \quad (2.65a)$$

and

$$\Pr [R_1|H_2] = \frac{e^{\frac{2\mu z}{\sigma^2}} - 1}{e^{\frac{2\mu a}{\sigma^2}} - 1}. \quad (2.65b)$$

These expressions match those found elsewhere in the literature for the mechanistic diffusion model's response probabilities. Importantly, although (2.61) assumes $E(0)$ corresponds to the true (objective) prior log-odds, changing the prior on H has no effect on $\Pr [R | H]$. Therefore (2.63), and hence (2.65), hold even if the starting point does not correspond to the objective prior. This fact should be obvious for the mechanistic model, because prior probabilities play no role in that model. Most

psychological applications of the diffusion model treat the starting point as a free parameter, not determined by the Bayesian analysis presented here.

In addition to response probability, the diffusion model predicts the joint distribution of response and RT (denoted here by the random variable T). These predictions are derived in numerous sources, and we simply repeat them here, first for the standard mechanistic parameterization:

$$\Pr [R_1, T \leq t|H_1] = \Pr [R_1|H_1] - \frac{\pi\sigma^2}{a^2} e^{-\frac{\mu(a-z)}{\sigma^2}} \times \sum_{k=1}^{\infty} \frac{2ka^2\sigma^2}{k^2\pi^2\sigma^4 + a^2\mu^2} \sin\left(\frac{k\pi(a-z)}{a}\right) e^{-\frac{k^2\pi^2\sigma^4 + a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.66a)$$

$$\Pr [R_2, T \leq t|H_1] = \Pr [R_2|H_1] - \frac{\pi\sigma^2}{a^2} e^{-\frac{\mu z}{\sigma^2}} \times \sum_{k=1}^{\infty} \frac{2ka^2\sigma^2}{k^2\pi^2\sigma^4 + a^2\mu^2} \sin\left(\frac{k\pi z}{a}\right) e^{-\frac{k^2\pi^2\sigma^4 + a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.66b)$$

$$\Pr [R_1, T \leq t|H_2] = \Pr [R_1|H_2] - \frac{\pi\sigma^2}{a^2} e^{-\frac{\mu(a-z)}{\sigma^2}} \times \sum_{k=1}^{\infty} \frac{2ka^2\sigma^2}{k^2\pi^2\sigma^4 + a^2\mu^2} \sin\left(\frac{k\pi(a-z)}{a}\right) e^{-\frac{k^2\pi^2\sigma^4 + a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.66c)$$

$$\Pr [R_2, T \leq t|H_2] = \Pr [R_2|H_2] - \frac{\pi\sigma^2}{a^2} e^{-\frac{\mu z}{\sigma^2}} \times \sum_{k=1}^{\infty} \frac{2ka^2\sigma^2}{k^2\pi^2\sigma^4 + a^2\mu^2} \sin\left(\frac{k\pi z}{a}\right) e^{-\frac{k^2\pi^2\sigma^4 + a^2\mu^2}{2a^2\sigma^2}t}. \quad (2.66d)$$

Translating these equations into the parameters of the Bayesian model yields

$$\Pr [R_1, T \leq t|H_1] = \Pr [R_1|H_1] - \frac{\pi\xi}{2\alpha^2} e^{-\frac{\alpha-E(0)}{2}} \times \sum_{k=1}^{\infty} \frac{4k\alpha^2}{k^2\pi^2\xi + \alpha^2\xi} \sin\left(\frac{k\pi(\alpha - E(0))}{2\alpha}\right) e^{-\frac{k^2\pi^2\xi + \alpha^2\xi}{4\alpha^2}t}, \quad (2.67a)$$

$$\Pr [R_2, T \leq t|H_1] = \Pr [R_2|H_1] - \frac{\pi\xi}{2\alpha^2} e^{-\frac{\alpha+E(0)}{2}} \times \sum_{k=1}^{\infty} \frac{4k\alpha^2}{k^2\pi^2\xi + \alpha^2\xi} \sin\left(\frac{k\pi(\alpha + E(0))}{2\alpha}\right) e^{-\frac{k^2\pi^2\xi + \alpha^2\xi}{4\alpha^2}t}, \quad (2.67b)$$

$$\Pr [R_1, T \leq t|H_2] = \Pr [R_1|H_2] - \frac{\pi\xi}{2\alpha^2} e^{-\frac{\alpha-E(0)}{2}} \times \sum_{k=1}^{\infty} \frac{4k\alpha^2}{k^2\pi^2\xi + \alpha^2\xi} \sin\left(\frac{k\pi(\alpha - E(0))}{2\alpha}\right) e^{-\frac{k^2\pi^2\xi + \alpha^2\xi}{4\alpha^2}t}, \quad (2.67c)$$

$$\Pr [R_2, T \leq t | H_2] = \Pr [R_2 | H_2] - \frac{\pi \xi}{2\alpha^2} e^{\frac{\alpha+E(0)}{2}} \times \\ \sum_{k=1}^{\infty} \frac{4k\alpha^2}{k^2\pi^2\xi + \alpha^2\xi} \sin \left(\frac{k\pi(\alpha + E(0))}{2\alpha} \right) e^{-\frac{k^2\pi^2\xi + \alpha^2\xi}{4\alpha^2}t}. \quad (2.67d)$$

For the remainder of this section we use the mechanistic parameterization, and derive some further properties of the model's predictions. Starting with the (cumulative) distribution functions in (2.66), we can take the derivative with respect to time to get the RT density functions,

$$p_i^j(t) = \frac{d}{dt} \Pr [R_j, T \leq t | H_i]. \quad (2.68)$$

These density functions are equal to

$$p_1^1(t) = \frac{\pi\sigma^2}{a^2} e^{\frac{\mu(a-z)}{\sigma^2}} \sum_{k=1}^{\infty} k \sin \left(\frac{k\pi(a-z)}{a} \right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.69a)$$

$$p_1^2(t) = \frac{\pi\sigma^2}{a^2} e^{-\frac{\mu z}{\sigma^2}} \sum_{k=1}^{\infty} k \sin \left(\frac{k\pi z}{a} \right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.69b)$$

$$p_2^1(t) = \frac{\pi\sigma^2}{a^2} e^{-\frac{\mu(a-z)}{\sigma^2}} \sum_{k=1}^{\infty} k \sin \left(\frac{k\pi(a-z)}{a} \right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.69c)$$

$$p_2^2(t) = \frac{\pi\sigma^2}{a^2} e^{\frac{\mu z}{\sigma^2}} \sum_{k=1}^{\infty} k \sin \left(\frac{k\pi z}{a} \right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}. \quad (2.69d)$$

The integral of each density function equals the corresponding total response probability:

$$\int_0^\infty p_i^j(t) dt = \Pr [R_j | H_i]. \quad (2.70)$$

Dividing by the response probability gives the conditional RT distribution for each response under each category,

$$q_i^j(t) = \frac{d}{dt} \Pr [T \leq t | H_i, R_j].$$

These conditional RT distributions are equal to

$$q_1^1(t) = \frac{\pi\sigma^2}{a^2} \cdot \frac{e^{\frac{\mu a}{\sigma^2}} - e^{-\frac{\mu a}{\sigma^2}}}{e^{\frac{\mu z}{\sigma^2}} - e^{-\frac{\mu z}{\sigma^2}}} \sum_{k=1}^{\infty} k \sin \left(\frac{k\pi(a-z)}{a} \right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.71a)$$

$$q_1^2(t) = \frac{\pi\sigma^2}{a^2} \cdot \frac{e^{\frac{\mu a}{\sigma^2}} - e^{-\frac{\mu a}{\sigma^2}}}{e^{\frac{\mu(a-z)}{\sigma^2}} - e^{-\frac{\mu(a-z)}{\sigma^2}}} \sum_{k=1}^{\infty} k \sin \left(\frac{k\pi z}{a} \right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.71b)$$

$$q_2^1(t) = \frac{\pi\sigma^2}{a^2} \cdot \frac{e^{\frac{\mu a}{\sigma^2}} - e^{-\frac{\mu a}{\sigma^2}}}{e^{\frac{\mu z}{\sigma^2}} - e^{-\frac{\mu z}{\sigma^2}}} \sum_{k=1}^{\infty} k \sin\left(\frac{k\pi(a-z)}{a}\right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}, \quad (2.71c)$$

$$q_2^2(t) = \frac{\pi\sigma^2}{a^2} \cdot \frac{e^{\frac{\mu a}{\sigma^2}} - e^{-\frac{\mu a}{\sigma^2}}}{e^{\frac{\mu(a-z)}{\sigma^2}} - e^{-\frac{\mu(a-z)}{\sigma^2}}} \sum_{k=1}^{\infty} k \sin\left(\frac{k\pi z}{a}\right) e^{-\frac{k^2\pi^2\sigma^4+a^2\mu^2}{2a^2\sigma^2}t}. \quad (2.71d)$$

Thus, we see that $q_1^j(t) = q_2^j(t)$, meaning the distribution of RTs for response j is the same regardless of whether that response is correct (Category $i = j$) or incorrect (Category $i \neq j$). All that differs between categories is the overall probability of that response. Moreover, if $z = a/2$ (i.e., no response bias, or equal priors for the two categories in the Bayesian model), then $q_i^1(t) = q_i^2(t)$. That is, for either stimulus category, the RTs for responses R_1 and R_2 have the same distribution (again, all that differs is the overall probability of the two responses). If the starting point is not midway between the thresholds, the model can predict faster RTs for one response than for the other. However, the direction of the effect will be the same for both categories: If RTs are faster for R_1 than for R_2 following Category 1 stimuli, they will also be faster for R_1 than R_2 following Category 2 stimuli.

2.8 Intertrial Variability and Unfalsifiability

What the diffusion model as presented thus far cannot predict (under either the mechanistic or Bayesian formulation) is a difference between conditional RT distributions for the two responses that depends on which response is correct. That is, it cannot predict a pattern wherein correct RTs are faster than error RTs (i.e., response R_1 faster than R_2 under Category 1, and vice versa under Category 2), and likewise it cannot predict a pattern of error RTs being faster than correct RTs. As it turns out, numerous experiments show these patterns, presenting an empirical challenge to the model.

One solution to this challenge, which has been widely adopted in the literature, is to extend the plain diffusion model as presented thus far by introducing intertrial variability in the model parameters. For example, one could assume that μ is a random variable, taking on different values on different trials of the same stimulus category. When μ is larger, the evidence rates for the two categories are better separated, and consequently the model responds more quickly and more accurately (for a fixed value of the threshold). Indeed, it is readily seen from (2.65) that increasing μ increases the probabilities of correct responses, $\Pr[R_1|H_1]$ and $\Pr[R_2|H_2]$, and calculations using (2.71) show that RTs become shorter as well. Likewise, decreasing μ increases the probabilities of error responses and produces longer RTs. When different values of μ are mixed across trials, the result is a variant of Simpson's paradox: even though errors and correct responses have the same RT distribution

for any given value of μ , errors are more likely on trials when μ is smaller, which is also when RT tends to be longer, and therefore errors are overall slower than correct responses once μ is integrated out.

Intertrial variability in the starting point can also break the symmetry between correct and error RT distributions, but in the opposite direction. Greater values of $E(0)$ or z produce faster RTs for R_1 , as can be calculated from (2.71a) and (2.71c), and they increase the probability of that response under either category, as can be seen from (2.63). However, the increase in response probability is greater under Category 2 (i.e., when R_1 is an error). That is, R_1 is more likely with starting points that lead to a short RT for that response, and this effect is more pronounced when the response is incorrect. The same conclusion holds for R_2 . Therefore, intertrial variability in the starting point produces error RT distributions that are faster than correct RT distributions.

Combining intertrial variability in the drift rate and starting point produces a more complex pattern. When the thresholds are relatively large, the impact of starting-point variability is reduced, and the drift-rate variability dominates to produce a pattern of slow errors. When the thresholds are relatively small, starting-point variability dominates to produce a pattern of fast errors. Experimental manipulations that are assumed to influence subjects' speed–accuracy trade-off, and that are modeled by changes in threshold (i.e., smaller thresholds under conditions encouraging speed, larger thresholds under conditions encouraging accuracy), have been found to yield this pattern. For example, when an instructional manipulation is used, errors tend to be faster than correct responses when subjects are told to emphasize speed over accuracy, and when subjects are told to emphasize accuracy this relationship tends to reverse.

The proposal of intertrial variability thus appears to be empirically successful, and to have resolved what is otherwise an important predictive failure of the diffusion model. However, this proposal suffers two problems. First, it is theoretically unmotivated. The general idea that sensory input is variable has a long history in psychophysics, but the diffusion model already incorporates this idea as within-trial variability. The proposal of two separate timescales of variability (within trials and between trials) seems to have been introduced solely to fit the data. The second problem, which is somewhat a consequence of the first, is that without some theory to constrain the form of the intertrial distributions, the model becomes excessively flexible. In fact, if the drift-rate distribution is entirely free, then the model becomes fully unfalsifiable. That is, for any joint distribution over the response and RT, there exist drift-rate distributions under which the model exactly reproduces that joint distribution.

To state this unfalsifiability result more formally: let $G_i^j(t)$ for $i,j \in \{1, 2\}$ be any set of non-decreasing right-continuous functions with $G_i^j(t) = 0$ and $\lim_{t \rightarrow \infty} G_i^1(t) + G_i^2(t) \leq 1$. (Allowing for inequality in the latter constraint allows a non-zero probability that no response is given.) Let a be any value for the upper threshold (with the lower threshold equal to 0), and let z be any value for the starting point (fixed across trials), with $0 < z < a$. Then there exist a value of σ and

intertrial drift-rate distributions under the two categories such that the diffusion model exactly predicts $\Pr[R_j, T \leq t | H_i] = G_i^j(t)$ for all $t \in \mathbb{R}^+$ and $i, j \in \{1, 2\}$.

To prove this statement, let $\sigma = 0$ and define the drift-rate distribution under each category by

$$\Pr[\mu \leq x | H_i] = \begin{cases} G_i^2(-\frac{z}{x}) & x < 0 \\ \lim_{t \rightarrow \infty} G_i^2(t) & x = 0 \\ 1 - \sup_{t < \frac{a-z}{x}} G_i^1(t) & x > 0. \end{cases} \quad (2.72)$$

Under the special case of no diffusion ($\sigma = 0$), the response is always R_1 if $\mu > 0$ and R_2 if $\mu < 0$, and the RT is given by

$$T = \begin{cases} -\frac{z}{\mu} & \mu < 0 \\ \infty & \mu = 0 \\ \frac{a-z}{\mu} & \mu > 0. \end{cases} \quad (2.73)$$

Therefore, for any $t > 0$ we have

$$\begin{aligned} \Pr[R_2, T \leq t | H_i] &= \Pr\left[\mu \leq -\frac{z}{t} \mid H_i\right] \\ &= G_i^2(t). \end{aligned} \quad (2.74a)$$

Likewise,

$$\begin{aligned} \Pr[R_1, T \leq t | H_i] &= \Pr\left[\mu \geq \frac{a-z}{t} \mid H_i\right] \\ &= 1 - \sup_{m < \frac{a-z}{t}} \left(1 - \sup_{\tau < \frac{a-z}{m}} G_i^1(\tau)\right) \\ &= \inf_{\frac{a-z}{m} > t} \left(\sup_{\tau < \frac{a-z}{m}} G_i^1(\tau)\right) \\ &= G_i^1(t) \end{aligned} \quad (2.74b)$$

with the last equality due to right-continuity of G_i^1 .

Although this proof relies on allowing diffusion to be absent from the model, one can also choose $\sigma > 0$ and obtain a model with predictions arbitrarily close to a given G_i^j . That is, for any $\epsilon < 0$, there exist a diffusion rate and drift-rate distributions such that the model's predictions satisfy

$$\left|\Pr[R_j, T \leq t | H_i] - G_i^j(t)\right| < \epsilon \quad (2.75)$$

for all $t > 0$ and $i, j \in \{1, 2\}$. This follows from the fact that (2.66) is continuous with respect to σ at $\sigma = 0$ and that probability functions are monotonic with compact range (thus ensuring uniform convergence in (2.75)). Because the model's predictions are invariant under any transformation of its parameters $(z, \mu, \sigma, a) \rightarrow (\gamma z, \gamma \mu, \gamma \sigma, \gamma a)$ for $\gamma > 0$, one can then pick a transformation that results in any

desired value of σ . In other words, if the diffusion rate is fixed in advance, one can still obtain a model with predictions arbitrarily close to a given G_i^j by appropriate choice of the drift-rate distributions, starting point, and thresholds.

In practical applications of the diffusion model, the intertrial distributions of drift rate and starting point are not fully unconstrained as they are in the proof just given. Instead, drift rate is typically assumed to vary according to a Gaussian distribution, and starting point according to a uniform one. This more restricted model is not unfalsifiable, and indeed it makes constrained predictions that have been well-supported empirically. However, the choices of Gaussian and uniform distributions are made purely for mathematical convenience; they are considered implementation assumptions rather than theoretical commitments. Therefore, we are left in an unusual situation, where a formal model makes constrained and successful predictions, but the theory this model is meant to embody (diffusion process, invariant boundaries, and intertrial variability in starting point and drift rate) is unfalsifiable. Clearly the model is capturing regularities in human decision-making behavior in a way that gives it remarkable predictive power, but at present the reasons for this empirical success are poorly understood.

2.9 Further Reading

For readers not familiar with Bayesian models of cognition, Griffiths, Kemp, and Tenenbaum (2009) provide a tutorial introduction.

The signal detection model was originated by Tanner and Swets (1954) and Green and Swets (1966), with important later elaborations by Ashby and Townsend (1986).

Reviews of evidence sampling models of speeded choice, including the random walk and diffusion models, can be found in Ratcliff and Smith (2004), Luce (1986), Townsend and Ashby (1983), and Vickers (1979).

The original formulation of the random walk model and its grounding in the SPRT are due to Stone (1960). Later developments can be found in Laming (1968), Link (1975), and Link and Heath (1975).

The Wiener diffusion model was originated by Ratcliff (1978). The Ornstein–Uhlenbeck (OU) model, which is closely related to the Wiener diffusion model but includes a decay component in the dynamics of the evidence process, was developed by Busemeyer and Townsend (1993).

Bogacz *et al.* (2006) present an optimality analysis of several evidence-accumulation models, including the diffusion and OU models, that is complementary to the Bayesian derivation presented here.

Response and RT predictions for the diffusion model can be found in Ratcliff (1978), with elaboration in Smith (2000). Predictions for the random walk model and relationships to the diffusion model's predictions can be found in Smith (1990).

Intertrial variability was first introduced by Laming (1968), who assumed a variable starting point for the random walk model. Ratcliff (1978) introduced a variable drift rate for the diffusion model. Ratcliff and Rouder (1998) assumed variability in both starting point and drift rate in the diffusion model and showed that these assumptions together can predict a crossover pattern of fast errors under speed instructions and slow errors under accuracy instructions. Data showing this crossover pattern can be found, for example, in Ratcliff, Van Zandt, and McKoon (1999), and Wagenmakers *et al.* (2008).

The unfalsifiability property of the diffusion model with unconstrained drift-rate distribution was proven by Jones and Dzhafarov (2014a). Further discussion of this theorem appears in Heathcote, Wagenmakers, and Brown (2014), Smith, Ratcliff, and McKoon (2014), and Jones and Dzhafarov (2014b).

2.10 Conclusions

We have shown here that the diffusion model that has been influential in psychological studies of speeded decision-making has a normative basis in Bayesian inference from a continuous evidence stream. The version of the model that results from this rational analysis is formally equivalent to the standard, mechanistic diffusion model and offers new insights on psychological interpretations of its parameters.

A general challenge to Bayesian models of cognition is that exact Bayesian inference becomes intractable as the task environment becomes more complex. For example, the rational analysis presented here does not cover cases where there are different stimulus subtypes within each category, or where the true drift rates of the input process (μ_i) are unknown, or where there are sequential dependencies across trials. Although extensions of the present model to these cases are possible, it seems that the brain must eventually give up exact Bayesian inference in favor of approximate methods. Therefore, as a cognitive theory, Bayesian optimality is better viewed as a guiding principle that is likely to be more accurate in simpler situations. In light of these considerations, we suggest the value of an analysis like the one presented here is that it offers a link between mechanistic and rational levels of explanation. Understanding the normative underpinnings of a mechanistic model, such as the diffusion model, may provide guidance in extending it to cover more complex tasks or phenomena.

One example of how a normative grounding might inform extensions of the diffusion model concerns the relationships between RT distributions for correct and error responses. We have reviewed here how the plain diffusion model makes strong predictions about these relationships that do not hold up empirically. The model can be extended to enable violations of these predictions by incorporating random intertrial variability in its drift rate and starting point, but this extension is atheoretical and comes at the cost of making the underlying theory unfalsifiable. In order to obtain a model that makes a genuine explanatory contribution – that

is, that makes theoretically driven, constrained predictions that match the data – it seems that what is needed is a theory of intertrial variability itself, one that implies constraints on the forms that variability can take. Because the Bayesian treatment presented here offers a rational interpretation of the starting point and drift rate, it may suggest a principled theory of how and why they vary. In particular, one hope is that such a theory could emerge from trial-by-trial learning of the task parameters (the prior probabilities of the categories, the drift criterion, and the signal-to-noise ratio) that the observer must know in order to carry out optimal inference.

References

- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154–179.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review*, 113, 700–765.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100, 432–459.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.
- Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In Ron Sun (ed.), *The Cambridge handbook of computational cognitive modeling*. Cambridge: Cambridge University Press.
- Heathcote, A., Wagenmakers, E.-J., & Brown, S. D. (2014). The falsifiability of actual decision-making models. *Psychological Review*, 121, 676–678.
- Jones, M., & Dzhafarov, E. N. (2014a). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121, 1–32.
- Jones, M., & Dzhafarov, E. N. (2014b). Analyzability, ad hoc restrictions, and excessive flexibility of evidence-accumulation models: Reply to two critical commentaries. *Psychological Review*, 121, 689–695.
- Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York, NY: Wiley.
- Link, S. W. (1975). The relative judgement theory of two choice response time. *Journal of Mathematical Psychology*, 12, 114–135.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77–105.
- Luce, R. D. (1986). *Response times*. Oxford: Oxford University Press.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 333–367.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Smith, P. L. (1990). A note on the distribution of response time for a random walk with Gaussian increments. *Journal of Mathematical Psychology*, 34, 445–459.

- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, 44, 408–463.
- Smith, P. L., Ratcliff, R., & McKoon, G. (2014). The diffusion model is not a deterministic growth model: Comment on Jones and Dzhafarov (2014). *Psychological Review*, 121, 679–688.
- Stone, M. (1960). Models for choice reaction time. *Psychometrika*, 25, 251–260.
- Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-409.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.
- Wagenmakers, E.-J., Ratcliff, R., Goméz, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, 58, 140–159.

3 Stochastic Foundations of Elementary Mental Architectures

Joseph W. Houpt, James T. Townsend, and Brett Jefferson

3.1	Introduction	104
3.2	General Event Spaces: Formal Probability Theory	109
3.3	Establishing Serial and Parallel Distinctions through the σ -Spaces	112
3.4	Causality in Parallel and Serial Systems	116
3.5	Experimental Identifiability of Architectural Distinctions of Causal Mental Systems	116
3.5.1	Distinguishing Parallel and Serial Systems with Selective Influence Manipulations	120
3.6	Discussion and Conclusions	125
	References	126

3.1 Introduction

In this introductory section, the language will be confined to prose and informality. Later, we will propose a novel general mathematical framework, within which our own as well as a number of other rigorous approaches may be nested.

At the most trite level, serial and parallel systems are different because in the former, sub-processes are taken up one at a time and only exactly when the previous sub-process is completed. In the latter, all sub-processes begin at the same instant, although they may finish at distinct points in time. However, this truly elemental distinction does not prevent mathematical descriptions of the two model classes from often being equivalent (see Townsend, 1972, 1974).

We will first follow the tactics of specific earlier papers as well as Townsend and Ashby (1983), especially Chapters 14 and 15, in delineating fundamental ways in which architectures can be similar and also when they can differ. Subsequently, a unifying and new foundational strategy will be pursued which unifies the homologues and heterologues under a single metatheatrical umbrella. With this strategy in place, we consider possibilities for empirical discriminability of serial and parallel systems or data that allows for model mimicry.

A “system” will be the actual concrete thing itself and of course could be realized in many ways from microchips to neurons. In order to keep the argot

reasonably minimal, we define a “model” to actually be any class of stochastic processes defined by axioms and/or sets of parameters.

Thus, an (admittedly simplistic) parallel system for recognizing pure tones might be a set of tuning forks displayed on a table, each one associated with a specific frequency. A serial rendition of the same kind of task might consist of a sequential exposure of each tuning fork in turn.

Systems possess “processors” which work on the objects fed to them. The “objects” may differ from one another in ways that might affect the time of a processor to complete its individual task. For instance, it has been proposed that the identity of an object could influence the processing time (Townsend, 1976; Snodgrass & Townsend, 1980; Townsend & Evans, 1983; Townsend & Ashby, 1983; Van Zandt & Townsend, 2012), with an important special case being match vs. mismatch of an object to a target item.

The objects to be processed are often associated with one or more dimensions of location. For instance, an auditory set of objects typically possesses a single dimension or order, whereas a visual presentation of letters is usually characterized by a two-dimensional spatial code.

We now turn to a list of informal properties that we associate with the common and distinctive properties of serial and parallel systems. There are two concepts that we use here that will require further formalization later. First, if the representation of an object might be changing in some meaningful way, we say that the object or process is “being worked on.” We give a more formal definition of “being worked on” in Definition 3.1. Second, a process is “finished” when there are either no additional changes that will occur in that process or, if additional changes do occur, they will have no effect on the system’s response. The formal definition of “finished” is given in Definition 3.3.

1. Properties Held in Common by Parallel and Serial Systems

- SP1** The physical properties of the objects can influence processing.
- SP2** The order in which previously completed processes finished can influence the processing of unfinished objects.
- SP3** The processing times can influence system completion times.
- SP4** The order in which the processes finish can influence system completion times.

2. Distinctive Properties of Serial Systems

- S1** At any point in time, at most one object is being worked on.
- S2** If the processing of an object has not yet started, the properties of that object cannot influence the processing of other objects.
- S3** A preordained order of process completions can influence all processes.

3. Distinctive Properties of Parallel Systems

- P1** At any point in time, any of the uncompleted objects may be being worked on.
- P2** Properties of unfinished objects can influence the processing of other objects through processing interactions.

P3 The order of unfinished process completions cannot influence those processes.

Even before formalizing these properties, it is clear that there are plenty of potential systems that could hold both distinctive properties of serial systems and the distinctive properties of parallel systems. For example, **P1** only allows that uncompleted objects may be worked on, but it does not require that they must be worked on, so a system that processes only one unfinished object at time would still satisfy this property. Hence, although it reduces the generality of the systems that may be described within this purview, we also consider a stronger version of **P1**.

P1* At any point in time, any of the uncompleted objects *must* be being worked on.

We will refer to parallel systems that satisfy **P1*** as *strict* parallel systems.

Alternatively, or additionally, we may also consider further constraint on the class of serial systems. For example, the properties outlined thus far allow for a serial system that can start processing an object, but then switch to another process before the first is completed. This would then allow for properties of unfinished objects to influence the processing of other objects, a property that we had nominally reserved for parallel systems. Thus, we consider this further constraint on serial systems,

S4 At any point in time, there is at most one process that has started but is not yet finished.

We will refer to serial systems that satisfy **S4** as *strict* serial systems. Note, along with **S4**, **S2** implies that unfinished objects cannot influence the processing of other objects.

Strict serial models can thus be defined as a distribution on processing order of the objects and a set of distributions for the processing of an object conditioned on that object being the one worked on at a given time (which may be dependent on processing order). Since only one object at a time is being processed, there is no “present status” of processes associated with other objects to depend on, outside of those completed in the past.

A parallel model can be defined as a distribution of processing where the ongoing processing of individual items can potentially depend on the status of any other process or object as well as the history to date of times and order of processing. It is barred from letting the processing order be influenced by subsequent/future order of processing.

We should note that a mental architecture designer, for instance nature, could trade off the ability of a serial system to make its distributions depend on a preordained order, with a dependence on historical events such as previous completion time. In point of fact, both pre-set order and the developing history of completion durations could be influential. Thus, the pre-set order could assign general distributions, such as first is Weibull, second is gamma, and so on. Then, the gamma rate parameter could be determined by the speed of the first stage, with a fast rate

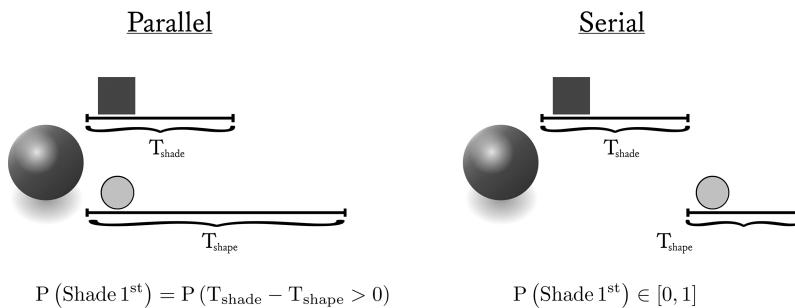


Figure 3.1 Illustration of a parallel and serial process identifying the shape and shade of an object. For the parallel process, both the shape and shade are worked on at the same time. In the serial process, the shape and shade are worked on at different times. In the parallel process, the probability of the order is a function of the completion time random variables, whereas the only constraint on the probability of the order for the serial process is that it is a probability (i.e., between 0 and 1).

if the first stage takes longer than 100 ms and a slow rate if it is faster than or equal to 100 ms.

The question of empirical identifiability of the foundational distinction of pre-determined order between parallel and serial processes will be discussed in a later section.

The causally acceptable ability of serial systems, but not parallel systems, to preselect the order of processing and processing time distributions (although not the actual realized processing durations) only requires a state space of positions of objects, but not of their composition, similarity to targets and so on. Examples are feature extraction and/or perceptual distance of memory object to a probe item (e.g., see Townsend, 1972). See Figure 3.1, which illustrates this parallel–serial disparity.

On the other hand, the latitude allowed parallel systems to permit various object processes to interact in ways so that states of processing of one (or more) process to depend on that of others calls for a finer-grained description of objects. For instance, feature extraction and computation of similarity to a probe item are examples (e.g., see Townsend, 1972, 1976). See Figure 3.2, which illustrates this parallel–serial disparity.

Our ensuing mathematical account will capture these notions more rigorously. Mathematical modeling departs from statistics and, for the most part, psychometrics, by way of the following precept (following Townsend & Ashby, 1983, chapter 15):

Principle of Correspondent Change

- Empirical changes in the environment of a stimulating situation should be reflected in a non-vacuous theory or model by corresponding changes or invariances in the model or theory.

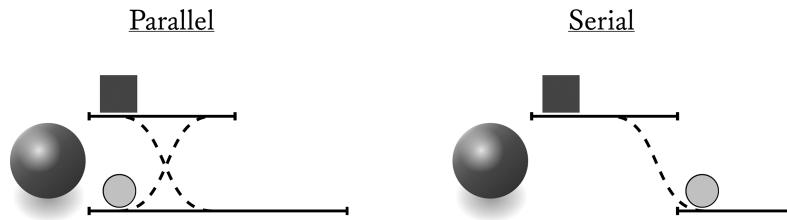


Figure 3.2 Illustration of interaction in parallel and serial systems. Currently processed information from one source can only affect present or future processing of other sources. Thus, while parallel systems can share information in both directions, serial systems can only share information from earlier processed items to later processed items.

- B. For any given empirical milieu and for any given class of models, there will exist a set of subclasses where models are indistinguishable within their subclass and in that specific milieu.

Principle of Selective Influence

The principle of selective influence (PSI) is a case of principle of correspondent change (PCC) and is likely the most valuable instance to date. The original informal concept was that environmental manipulations can be found which, realized stochastically, slow down or speed up two or more sub-processes (e.g., Sternberg, 1969). Initially, predictions were confined to mean RTs of serial systems, although speculations were made about parallel systems. Additionally, Sternberg (1973) suggested analyses of variance and higher order cumulants. Later, mean RT predictions were offered for more complex systems (e.g., Schweickert, 1978). It also came to be realized that assumptions about how an experimental factor affected an object's response time (RT) distribution were vital to proving claims about parallel vs. serial and more complex architectures, even at the level of mean RTs (Townsend & Schweickert, 1989; Townsend, 1984, 1990; Townsend & Thomas, 1994). Theoretically deep aspects of PSI have been discovered by Dzhafarov and colleagues in recent years (Dzhafarov, 2003; Dzhafarov, Schweickert, & Sung, 2004; Dzhafarov & Gluhovsky, 2006; Kujala & Dzhafarov, 2008; Dzhafarov & Kujala, 2010; Zhang & Dzhafarov, 2015). The concept of selective influence has been generalized by them and others (e.g., Schweickert, Fisher, & Sung, 2012; Algom *et al.*, 2015) to include other dependent variables such as accuracy. The valuable construct of selective influence will be further discussed subsequently.

The PCC is broader still, however, as can be seen in the contrast between the causal confinement of parallel processes to the stochastic processes, dependence on evolving history vs. the ability of a serial system to causally be a function of such future orders. Such a fundamental distinction could, in principle, be associated with observable differences in behavior, whether or not the investigator exploits these differences via manipulation of selective factors.

3.2 General Event Spaces: Formal Probability Theory

The fundamental concept of formal probability theory is that of a probability space. That space is given by a set of elements, usually denoted by Ω , specified subsets of the full space for which a probability is defined, \mathcal{F} , and the probability associated with each of those subsets, P . Those triples (Ω, \mathcal{F}, P) that satisfy the constraints we summarize below are a probability space (also referred to as a probability triple). In light of our discussion above, Ω could be the possible completion times of each of the processes, $(\mathbb{R}^+)^n$ for n processes. Alternatively, Ω could be the possible states of each of the processes at any time, $\mathbb{R}^n \times \mathbb{R}^+$, where each of the dimensions of \mathbb{R}^n corresponds to the activation state of a process and \mathbb{R}^+ corresponds to time since stimulus onset.

The subsets in \mathcal{F} can include any subsets of Ω that would normally be useful for distinguishing serial and parallel processing, but not all subsets of Ω are required to be in \mathcal{F} . \mathcal{F} must include the entire space Ω and its complement, the empty set \emptyset . Also, for any set $A \in \mathcal{F}$, the complement A^c must be in \mathcal{F} . \mathcal{F} must also be closed under countable unions and countable intersections. Thus, if A_1, A_2, \dots are in \mathcal{F} , then $\cup_{i=1,2,\dots} A_i$ and $\cap_{i=1,2,\dots} A_i$ must also be in \mathcal{F} . When a collection of sets satisfies these constraints it is called a σ -algebra or σ -field.

For modeling cognitive processing times, we want to be certain that intervals are included in \mathcal{F} . For example, we will need to refer to probabilities such as “the probability that process A takes less than one second and process B takes between two and three seconds.” The smallest σ -algebra on the real numbers that includes all intervals is called the Borel σ -algebra. That this σ -algebra exists follows from the famous Kolmogorov extension theorem (e.g., Billingsley, 1995, §3).

For P to be a probability measure, the probability of the empty set must be zero, $P(\emptyset) = 0$, the probability of the whole space must be one, $P(\Omega) = 1$, and P must be countably additive. Countable additivity means that for any *disjoint* sets, A_1, A_2, \dots in \mathcal{F} , the probability of the union of the sets (which must be in \mathcal{F}) is the sum of the probability of each set, $P(\cup_{i=1,2,\dots} A_i) = \sum_{i=1,2,\dots} P(A_i)$.

In general, any probability measure that satisfies these conditions could be used for modeling mental processing. In practice, models tend to take one of two forms. Many modelers will specify the probability measure by associating the random events with commonly used distributions, such as an exponential distribution or a more complex distribution based on processing assumptions, such as the linear ballistic accumulator (Brown & Heathcote, 2008) or the drift-diffusion model (Ratcliff, 1978). Alternatively, one can study systems by determining what inferences can be made about the systems using only minimal additional constraints on the probability measure (c.f. Townsend & Nozawa, 1995; Dzhafarov *et al.*, 2004; Zhang & Dzhafarov, 2015).

We mentioned random variables and distributions, with which the reader will no doubt be at least minimally familiar, but these terms also have a more rigorous definition within probability theory. A univariate random variable X is a measurable function from the space Ω to the real numbers. To be measurable, the preimage of

every Borel set under X must be in the σ -algebra, \mathcal{F} , on Ω , i.e., for any Borel set B , $X^{-1}(B) \in \mathcal{F}$.

Suppose, for example, that instead of working with a space Ω that is defined on the response times, which are elements of \mathbb{R} , Ω is the space of a participant's actions in an experiment. Then the elements of Ω are not real numbers, but actions. To discuss response times, we would need to map from the space of actions to the real number representing the time it takes for an action. The measurability requirement allows one to work with probabilities of either the response time or the actions because there are connections between the measurable sets \mathcal{F} in Ω and measurable (specifically Borel) sets in \mathbb{R} .

The distribution, μ , of a random variable X is the function on the Borel sets that gives the probability measure of the corresponding set in \mathcal{F} given by X . Thus, if B is a Borel set, then $\mu(B) = P(X^{-1}(B)) = P(X \in B)$ where P is the probability measure from the space on which X is defined.

A common practice in RT modeling is to assume that response times have a familiar distribution, such as a gamma or Weibull distribution. This approach sets the measure on the Borel sets on the real line rather than setting the measure on a separate probability space of interest. In many cases, researchers are only interested in direct statements about the response time distributions, so there is no loss in modeling them directly with a distribution. As we see below, there is some danger in ignoring the complexity of the processes that lead to the response time when considering the effect of experimental variables on the response time distributions.

A single random variable can be quickly generalized to a multivariate random variable or random vector. If X is an n -dimensional random vector, then each dimension is a random variable on the same probability space. That is, each X_i is a measurable function from the probability space to the real numbers. The random vector has a probability distribution given by $\mu(B) = P(X \in B) = P(X_1 \in B_1, \dots, X_n \in B_n)$ when B is an n -dimensional Borel set. This measure μ is the joint distribution of the random variables X_1, \dots, X_n . Because each X_i is a random variable, we also have the marginal distributions, $\mu_i(B_i) = P(X_i \in B_i)$.

From multivariate random variables, we can generalize to stochastic processes. These are essentially multivariate random variables, but with a possibly infinite, even continuous, index set. Formally, a stochastic process can be defined by the set of random variables $\{X_t\}_{t \in T}$ on a probability space (Ω, \mathcal{F}, P) , where T is a possibly infinite index set, such that for any finite subset $A \subset T$, $\{X_t\}_{t \in A}$ is a multivariate random variable.¹

Most commonly, a stochastic process is a process that unfolds over time so we may be interested in the future values, $\{X(t), t > s\}$, given the value at a specific time, $X(s)$. In this case, we know more about the random variables as time progresses. For example, if we were using a stochastic process to model the movement

¹ To be able to add constraints to stochastic processes that depend on an uncountable subset $S \subset \{X(t)\}_T$, we need a more complex definition. While these constraints are important (e.g., continuity of the sample paths), we avoid the additional detail here for the sake of brevity.

of a mouse cursor across the screen, then at a given time t we would know where the mouse had been at every time up until t , i.e., if $X(s)$ is the cursor position at time s , then we could have observed every $X(s)$ for $s \leq t$. The formal way of describing this increase in knowledge over time is by specifying a collection of σ -fields that increase over time: for all t there is a $\mathcal{F}_t \in \mathcal{F}$ such that for any $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$. This type of collection of σ -fields is known as a filtration. We can capture an increase in knowledge about a stochastic process over time by requiring that the random variables $X(t)$ are each \mathcal{F}_t measurable where the collection of \mathcal{F}_t is a filtration.

Stating that $X(t)$ is measurable \mathcal{F}_t means that the possible events of the system \mathbf{X} at time t are included in the σ -field \mathcal{F}_t . Given a system \mathbf{X} , there is a naturally induced filtration constructed as the union of the smallest σ -fields for each t ,

$$\mathcal{F}_t = \bigcup_{s \in [0, t]} \left\{ X(s)^{-1}(B) : B \text{ is a Borel set} \right\}.$$

For the induced filtration, it is clear that given \mathcal{F}_t we have complete knowledge about $X(t)$, however, we do not necessarily know anything about $X(t')$ for $t' > t$.

Example 3.1 Consider a single process $X(t)$ defined at discrete points in time $t \in \mathbb{N}$. For each t , let $X(t)$ be continuous random variables defined on \mathbb{R} . Then each \mathcal{F}_i , where $i \in \mathbb{N}$, can be thought of as the disjoint union of i copies of Borel sets. So one event of \mathcal{F}_2 , for example, is the disjoint union $(2, 26) \sqcup (1, 5.56)$. An event of \mathcal{F}_{40} could be the disjoint union of 40 copies of $(0, 1)$!

Note two things from this example. First, the notion of a filtration is separate from the individual probabilities associated with each $X(t)$, although the example makes no mention of any specific distribution. Second, we could define $X(41)$ to be a discrete random variable. Doing so would make it clear that \mathcal{F}_{40} does not imply anything about the subsets/events of $\{X(41)^{-1}\}$. Generally, it is useful to think about each $\{X_i(t)\}$, as independent continuous random variables.

The next important concept is of a conditional probability. Suppose we have random variables X and Y on the measure space (Ω, \mathcal{F}, P) . In many cases we want to examine the properties of X given knowledge of Y (for example, the response times from trials when a person is correct). As long as the event we are conditioning on has positive probability, then the traditional definition is sufficient, $P(X|Y) = P(X, Y)/P(Y)$. However, if $P(Y)$ is zero, then this would suggest that $P(X|Y)$ is undefined or infinite. That might work if you were conditioning on correct or incorrect when there is no chance of the participant being incorrect because we would never worry about conditioning on something that is impossible. But what about conditioning on a particular response time? If a response time could be any positive real number (or even any positive real number less than some upper bound) then the probability that it is a specific value would have to be zero for most values. Nonetheless, the response time must be some value. Suppose that we observe a response time of 326 ms. In the conditional $P(\text{Correct} | \text{RT} = 326 \text{ ms})$ the probability of $\text{RT} = 326 \text{ ms}$ is zero.

To deal with the issue of conditioning on measure zero events, conditional probabilities are defined as a function of the conditioned variable, $f(B) = P(Y|X \in B)$. For this function to make sense, it needs to satisfy two requirements. First, the function needs to be measurable with respect to X (actually $\sigma(X)$: the σ -algebra generated by X). Second, we want to preserve the equality $P(X, Y) = P(X)P(Y|X)$. Formally, we require that for all $B \in \sigma(X)$, $\int_B P(Y|X \in B) dP = P(Y, X \in B)$ where the dP refers to the measure P on the probability space.

The next concept from the theory of stochastic processes that we will need is that of a stopping time. A stopping time is a random $\tau \in T$ such that $\{\tau \leq t\} \in \mathcal{F}_t$ for all t in the index set T . For example the stopping time could be a fixed time, e.g., $\tau = 326$ ms, or it could be based on some event that is measurable with respect to \mathcal{F} , e.g., τ is the first time that $X_i(t) \geq 10$. For those familiar with information accumulator models, they often assume information is accumulated up to some threshold, then a response is made. In this case the response time is a stopping time for the σ -field generated by the amount of information accumulated at each time t . When discussing the exhaustive and first-terminating stopping rules for mental processes in the next section, we will make use of the fact that if τ and ν are both stopping times for a given σ -field, then so are the minimum and the maximum of τ and ν .

In addition to the stopping times for processes, we will also be interested in the order in which processes complete. For example, if the processes X_1, X_2 and X_3 complete at T_1, T_2 and T_3 , respectively, we may be only interested in whether $T_2 < T_1 < T_3$ or $T_1 < T_2 < T_3$, etc. To refer to the first process to complete, when we are not concerned with which process it is, we will use $X_{(1)}$ and similarly $T_{(1)}$ for the completion time of the first process to finish and $t_{(1)}$ for a particular observed value of the completion time of the first process to be finished.

With these fundamental concepts in mind, we now turn to a formalization of the elementary properties of mental processes outlined in the introduction, using stochastic processes.

3.3 Establishing Serial and Parallel Distinctions through the σ -Spaces

We begin by formalizing the properties of serial and parallel systems that we outlined above. To do so, we connect the informal idea of “processes” developed above with the mathematical concept of adapted stochastic processes. If there are n elements of information that the cognitive system is working with, then we need a stochastic process corresponding to each of those elements, $\{X_i(t), t \in [0, \infty)\}_{i=1 \dots n}$. Equivalently, we can represent these processes as an n -dimensional stochastic process $\mathbf{X}(t), t \in [0, \infty)$. In this case, we can call each X_i a sub-process. We also want this process to be adapted to preserve causality. Our main goal is that future events do not change the present (i.e., $\mathbf{X}(t)$ is determined by events in the past $s < t$ or in the present, but not the future $s > t$). For this goal,

it is sufficient that t is ordered in the sense that $\mathbf{X}(t)$ is measurable with respect to the filtration \mathcal{F}_t . To formalize the property that a serial system may have a predetermined order of processing, we allow that the order is measurable \mathcal{F}_0 , i.e., that the order might be set at $t = 0$ or some $t < 0$.

Example 3.2 For concreteness, suppose a person is asked to judge the size, weight, and color of a ball. A reasonable assumption is that for each dimension of the stimulus, the associated sub-process has the form of some information accrual process such as $X_i(t) \stackrel{d}{\sim} \text{unif}(t_{i_1}, t_{i_2})$ for each t_{i_1} and t_{i_2} functions of t . The simplest case is to allow the $t_{i_1} = 0$ and $t_{i_2} = t$. Here, as time passes, the probability of acquiring more information about each dimension grows linearly. Naturally, we could impose some dependence across sub-processes, but for illustrative purposes consider the independent case. We will refer to this example as needed throughout the chapter.

The first distinguishing property of serial and parallel systems concerns “being worked on.” We can formalize “being worked on” as follows:

Definition 3.1 Process i is being worked on at time t if there is no $s > t$ such that for all $v \in (t, s]$,

$$P(X_i(t) = X_i(v)) = 1.$$

Less formally, for any amount of time that passes, there is a chance that X_i changes state during that time. Similarly, if the process is not being worked on, there is some amount of change in time for which there is no chance that the state will change. If color processing is delayed as above, then color is not being worked on for all $t \in [0, T_1)$. Note that, according to this definition, if X_i changes state at deterministic intervals, then X_i is not being worked on during the interval. This may seem odd for a discrete time system, but this can be mitigated by assuming t, s , and v are positive integers in the definition.

Next, we need definitions for starting and completing processing.

Definition 3.2 A sub-process $X_i(t)$ has started if for any $s \leq t$, X_i was being worked on at s . When the minimum (or infimum) of such s exists, s' , we say that X_i started at s' .

In Example 3.2, we could delay processing of color until size has completed. This would be realized as

$$X_3(t) \stackrel{d}{\sim} \begin{cases} 0 & \text{if } t \leq T_1 \\ \text{unif}(0, t) & \text{if } t > T_1. \end{cases}$$

The processing of color hasn’t started until time T_1 .

To formalize process completion, we use the first passage (or first exit) time concept.

Definition 3.3 Let each sub-process $X_i(t)$ have a predetermined completion region, $\theta_i \subset [0, \infty)$. We will say that $X_i(t)$ is complete (or finished) if $t \geq \tau =$

$\min_s X_i(s) \in \theta_i$. For the system completion time, we have the completion region $\Theta \subset \mathbb{R}_+^n$ and say that the system has completed if $t \geq \min_t \mathbf{X}(t) \in \Theta$. A system may need all sub-processes to be complete (exhaustive system), or it may need only a subset of the processes to be complete (referred to as a self-terminating system).

Example 3.3 From Definition 3.3, we can consider the three sub-process system $\mathbf{X} = \{X_1, X_2, X_3\}$ described above with completion regions $[1, 2]$, $[3, 4]$, and $[5, 6]$, respectively. An exhaustive system can be constructed by defining $\Theta = [1, 2] \times [3, 4] \times [5, 6]$. All sub-processes must be in their respective completion regions simultaneously in order for processing to complete. Likewise, a self-terminating system can be constructed by defining

$$\begin{aligned}\Theta = & [1, 2] \times (-\infty, \infty) \times (-\infty, \infty) \\ & \cup (-\infty, \infty) \times [3, 4] \times (-\infty, \infty) \\ & \cup (-\infty, \infty) \times (-\infty, \infty) \times [5, 6],\end{aligned}$$

the union of infinite strips. Here any sub-process may enter its respective completion region for processing to complete.

One may wonder what happens after completion. Perhaps the simplest assumption would be that processing stops, i.e., if τ_i is the completion time of X_i , then for all $t > \tau_i$, $X_i(t) = X_i(\tau_i)$. This has the advantage of connecting channel completion times to system completion times with simple rules (e.g., the system completion time of an exhaustive system is the maximum of the sub-process completion times). In some cases, requiring processing to stop upon completion leads to more constraints, which may be unintuitive. For example, it precludes the possibility that a participant might second-guess himself after responding. At the other end of the spectrum, we could leave the process totally unconstrained, free to move in and out of the completion region and free to have any continued processing influence unfinished processes. For convenience, we take the middle ground and assume that, although processing may continue after completion, the outcome of the (sub-) process does not change and changes in sub-processes after they have completed cannot influence other processes.

We now augment the framework to codify the properties that we outlined that both serial and parallel systems should satisfy. We follow Dzhafarov (2003) by limiting the probability space associated with a perceptual process to the internal states of the system and assume that the environment in which the system operates is deterministic. This assumption is not necessarily required for our theory, but it will simplify the connection between these systems and the theory of selective influence later in the chapter. Hence, we identify the probability space with a particular configuration of the environment, i.e., if ζ indicates the specifics of the environment, then the relevant probability space is $(\Omega_\zeta, \mathcal{F}_\zeta, P_\zeta)$. This allows for each of the processes defined on the space to be different under different environmental conditions, and in particular, satisfies **SP1**. In fact, this formalization

requires that the processes are different under different configurations of the environment because equality (and almost-sure equality) in measure theory are defined with respect to the underlying measure space. It may be possible that there are two different probability spaces that nonetheless lead to exactly the same distributions for \mathbf{X}_t for all t even when the two event spaces are not equivalent. Hence, in keeping with **SP1**, we also require that the environmental conditions are allowed to lead to different distributions on \mathbf{X}_t .

Property SP1 Let ζ_a and ζ_b be any two possible states of the world. Then it is possible that $(\Omega, \mathcal{F}_{\zeta_a}, P_{\zeta_a}) \neq (\Omega, \mathcal{F}_{\zeta_b}, P_{\zeta_b})$.

Property SP2 Let $O(I)$ be the order of T_i for $i \in I$. If $I_t = \{i : T_i < t\}$ then $\sigma(O(I_t)) \subset \mathcal{F}(t)$.

Property SP3 For all $i, \sigma(T_i) \in \mathcal{F}_T$.

Property SP4 Let $O(I)$ be the order of T_i for $i \in I$. If I is the set of all sub-processes, then $\sigma(O(I)) \subset \mathcal{F}_T$.

S1 states that a serial system can only have one process being worked on at any given time. Formally, this can be stated as follows.

Property S1 In a serial system, if, for any i , there is no $s > t$ such that for all $v \in (t, s]$, $P(X_i(t) = X_i(v)) = 1$ then for all $j \neq i$, there exists some $s > t$ such that for all $v \in (t, s]$, $P(X_j(t) = X_j(v)) = 1$ (or, equivalently, $P(X_j(t) \neq X_j(v)) = 0$).

Property S2 Let ζ_a and ζ_b be any two possible states of the world such that for any process i that has been worked on by t , object i is the same in both ζ_a and ζ_b . Then, in a serial system, $\mathcal{F}_{t, \zeta_a} = \mathcal{F}_{t, \zeta_b}$ and $X_a(t)$ defined on $(\Omega, \mathcal{F}_t, \Pr_{\zeta_a})$ is equal in distribution to $X_b(t)$ defined on $(\Omega, \mathcal{F}_t, \Pr_{\zeta_b})$.

Property S3 Let $O(I)$ be the order of T_i for $i \in I$. In a serial system, if I is the set of all sub-processes and for all $t > 0$, it is possible that $\sigma(O(I)) \subset \mathcal{F}(t)$.

Consider again Example 3.1. Now suppose that the order of serial processing was always color first. Then \mathcal{F}_0 would contain two events,

$$\mathcal{F}_0 = \{\{\text{color first, size second, weight third}\}, \{\text{color first, size third, weight second}\}\}.$$

Property S4 In serial systems, for all t , if there exists i such that $0 < X_i(t) < \theta_i$, then for all $j \neq i$, either $X_j(t) = 0$ or $T_j < t$.

Property P1 In parallel systems, for all t , and all i such that $T_i > t$, it is possible that there is no $s_i > t$ such that for all $v \in (t, s_i]$, $P(X_i(t) = X_i(v)) = 1$.

Property P1* In strict parallel systems, for all t , and all i such that $T_i > t$, there is no $s_i > t$ such that for all $v \in (t, s_i]$, $P(X_i(t) = X_i(v)) = 1$.

Property P2 In parallel systems, for any object i under the state of the world ζ_a such that $T_i > t$, if there exists another object j such that $T_j > t$, and another state of the world ζ_b in which there is an object j_b identical to object j from ζ_a , then it is

possible that $\mathcal{F}_{t,\zeta_a} \neq \mathcal{F}_{t,\zeta_b}$ and in particular that $X_j(t)$ defined on $(\Omega, \mathcal{F}_t, \Pr_{\zeta_a})$ may not be equal in distribution to $X_{jb}(t)$ defined on $(\Omega, \mathcal{F}_t, \Pr_{\zeta_b})$.

Property P3 Let $O(I)$ be the order of T_i for $i \in I$. In parallel systems, if $I = \{i : T_i > t\}$, then $\sigma(O(I)) \not\subseteq \mathcal{F}(t)$.

When **S4** is assumed, i.e., strict serial systems, then **S1** implies that the identities of unfinished objects that are not being worked on cannot influence the processing of other objects. Here, we use the filtration concept to encode the idea of what has influence on an outcome. If object i is unfinished and not being worked on at time t , then the processing of all objects $j \neq i$ at time t cannot be influenced by object i .

Thus any set in \mathcal{F}_t that one could condition upon when measuring the probabilities associated with X_j contains no information about X_i . In parallel systems, it is possible for information about unfinished processes to influence other processes, so there is no such constraint on the filtrations.

3.4 Causality in Parallel and Serial Systems

Although the current definitions are similar in many ways to the definitions presented in Townsend and Ashby (1983), the use of filtrations and adaptations are new. Townsend and Ashby (1983) defined models of parallel and serial systems as probability triples associated with the sample space of subprocesses' completion times and orders, so they did not investigate the notion of differences in the filtrations associated with parallel and serial models. We are engaging the notion of filtrations to achieve our goals of more formally constraining the definitions of parallel and serial systems with notions of causality.

In particular, for the models to maintain standard notions of causality, future events should have no influence on past events. For example, the total system completion time (i.e., exhaustive processing of all objects) should not influence the completion time of the first of many sub-processes to complete. In terms of adapted stochastic processes, the overall system completion time is not in the adapted σ -algebra at the first sub-process's completion time. For example, if there are three sub-processes but only the first has finished ($T_1 < t < T_2$), then the system completion time is not measurable \mathcal{F}_t , i.e., it does not make sense to talk about probabilities such as $P(T \in [a, b])$ or $P(T \in [a, n] | T_1 = \tau)$. However, if $t \geq T$, then both the system completion time and the first sub-process completion time are in the filtration \mathcal{F}_t , which allows us to consider probabilities such as $P(T \in [a, b] | T_1 = \tau)$ or $P(T_1 | T = s)$.

3.5 Experimental Identifiability of Architectural Distinctions of Causal Mental Systems

We now turn our attention to focus on the ways in which the underlying event spaces (i.e., the σ -fields) contrast between serial and parallel models. These distinctions differ in their empirical consequences. Some lead to predictable

and potentially testable differences even within very elementary experimental conditions, whereas others require more subtle treatments.

The distinguishing characteristics given in the introduction and formalized in Section 3.3 are not alone sufficient for the systems to always be distinct, because each of the constraints on one system (e.g., that serial systems can have at most one process being worked on) is only stated as something the other is allowed, but not required, to do (e.g., in a parallel system, any unfinished process being worked on). This leads to sufficient conditions for rejecting one class of system or the other, but in cases where the constraint is satisfied, we are not able to distinguish between the models. For this reason, we have also introduced a strict version of **P1** to allow for distinguishing between the two classes of systems.

First, suppose we can observe the filtration \mathcal{F}_t for all t at which a process completes. As long as at least two sub-processes have not yet finished, then a serial system is allowed to have more information about the order of processing in each of those \mathcal{F}_t than a parallel system. For example, if the serial system determines the complete order immediately once processing begins, then the order is in \mathcal{F}_t for all $t > 0$. Because a parallel system cannot predetermine the order of processing, then whenever there are still two unfinished processes, their order is not yet in \mathcal{F}_t . Once all but one of the processes has completed, that unfinished process will be the last, so the order of processing is determined (i.e., in \mathcal{F}_t). Because there is no requirement that serial systems *must* determine the full order before processing begins, this is merely a sufficient condition for rejecting parallel systems.

Suppose instead that we have access to the complete distribution over the processes when the system is finished, i.e., the distribution of $\mathbf{X}(s)$ for all s up to some t greater than or equal to the total completion time of the system. In terms of filtrations, we have full knowledge of the distributions of random variables \mathbf{X} that are measurable \mathcal{F}_t for some t greater or equal to the system completion time. We want to emphasize that this information is about the probabilities associated with different states of the system, not about particular sample paths that the system actually follows. At this point the order of completion is measurable, whether the system is parallel or serial, and hence the presence of order in the σ -field is no longer a distinguishing factor. Nonetheless, based on this strong assumption, we could easily distinguish between strict parallel and serial systems by conditioning on the event that a particular object is being worked on at a given time and check if any other processes are being worked on.

If, instead of accessing properties of the filtration and/or distributions, we only have access to sample paths of the system, we can check sample path properties to potentially reject serial systems.

Proposition 3.1 *The derivatives of the sample path of at most one sub-process from a serial system can be either be non-zero or non-existent at any $t > 0$.*

Proof Consider a sample path x_i of X_i that is not being worked on at time t . By the definition being worked on, there exists some $s_i > t$ such that for all $v \in (t, s]i$, $x_i(v) - x_i(t) = 0$ almost surely. Hence, for any $v \in (t, s)$, the derivative

of $x_i(v)$ exists and is zero almost surely. Now consider $s = \min_i s_i$ for all x_i that are not being worked on at t (which exists as we are assuming a finite number of sub-processes). Then the derivatives of the sample path of each of those sub-processes exists and is zero for some interval after t . Because, in a serial system, there is at most one sub-process being worked on at any t , there is an interval after any t for which at most one sub-process's sample path is either non-zero or non-existent. \square

In contrast, in a strict parallel system, regardless of the interaction among processes, there is some possibility that all of the unfinished processes change within an arbitrarily small time interval of each other.

Proposition 3.2 *In a strict parallel system, there is positive probability that for some $t > 0$ and any $\epsilon > 0$,*

$$X_1(t + \epsilon) - X_1(t) > 0 \text{ and } X_2(t + \epsilon) - X_2(t) > 0.$$

Proof Strict parallel systems must have all processes either being worked on or finished at all times, so for any time $t > 0$ and $\epsilon > 0$, if X_i are the sub-processes that have not yet finished, the probability $P[X_i(t + \epsilon) - X_i(t) = 0] \neq 1$. Hence, for all i , $P[X_i(t + \epsilon) - X_i(t) \neq 0] > 0$. \square

In practice, we may observe a sample trajectory on which no sub-processes change within a given ϵ of one another. Nonetheless, if we continue to observe more sample trajectories, we would eventually observe one or more processes changing within the same ϵ interval, no matter how small the ϵ .

Next, we show that parallel and serial systems cannot be distinguished based on the system completion times under a single experimental condition. Recall that the system completion time distribution is only meaningful for \mathcal{F}_t in which t is larger than system completion time.

Proposition 3.3 *Suppose we only have access to the distribution over completion time (and hence also order) of the sub-processes in the system.² Without further restrictions, serial system and strict parallel systems cannot be distinguished at this level.*

Proof Consider an arbitrary serial process with distributions $P(\mathcal{O})$ on the order of processing and $f(t_1, t_2, \dots, t_n | \mathcal{O})$ on the time it takes to complete each sub-process. Following Townsend and Ashby's (1983) notation, we let the random vector (A_1, A_2, \dots, A_n) be the vector in which each element gives the ordinal rank in which that item finished. Then a parallel process can completely mimic the serial distribution, f , by setting the completion time distribution of each object in the parallel system equal to the sum of the completion time of each object that completed before it along with the objects' completion time in the serial system. That is, for every possible outcome $\omega \in \Omega$,

² This is the same case that Townsend and Ashby (1983) examine in Chapter 14.

$$T_i^p(\omega) = T_i^s + \sum_{j: A_j(\omega) < A_i(\omega)} T_j^s.$$

In particular, this means that the probability over orderings are the same for both models.

We can use a similar mapping in reverse to arrive at a parallel system starting from a serial system,

$$T_i^s(\omega) = T_i^p(\omega) - T_{j:a_i-a_j=1}(\omega).$$

Under this reverse mapping, the density over orderings is given by

$$p = \int_0^\infty \int_{t_{(1)}}^\infty \cdots \int_{t_{(n)}}^\infty g(t_{(1)}, t_{(2)}, \dots, t_{(n)}) d\tau_{(n)} \dots d\tau_{(1)}.$$

This mapping is quite general, so if the only observables are the intercompletion times T (and hence the orderings), serial and parallel systems can be nearly non-discriminable. Consider the system at the completion of the first process, $t_{(1)}$. Under a serial system, the entire order of completion may be in the measure space at $t_{(1)}$ and the ordering of later processes can influence $t_{(1)}$. This is, of course, impossible in a parallel system. Unfortunately, if we cannot observe the ordering at $t_{(1)}$, then there is no way to test whether future completion orderings have an effect on $t_{(1)}$. Once we have observed the completion order, we could evaluate the conditional probabilities,

$$P(\tau_1 \in \mathcal{T} | \langle A, B, C \rangle) \stackrel{?}{=} P(\tau_1 \in \mathcal{T} | \langle A, C, B \rangle).$$

However, by the above mapping, whichever pattern of dependence we observe in a serial system can be mimicked by a parallel system. \square

Although the mapping holds in general, it can lead to some unintuitive consequences. First, for a serial system, the probability of process a finishing first is the sum of the density when a is first under order $O_1 = a, b, c$ plus the density when a is first under order $O_2 = a, c, b$. If we assume within-stage independence in the parallel system (a system is within-stage independent if, for any interval of time during which no sub-process completes, all sub-processes that have not yet finished are independent on that interval), then for a parallel system to perfectly mimic this system,

$$\begin{aligned} & P(\mathcal{O}_1)f_{a_1}(t_{a_1}|\mathcal{O}_1) + P(\mathcal{O}_2)f_{a_1}(t_{a_1}|\mathcal{O}_2) \\ &= P(T_{a_1} = t_{a_1} < T_{b_1}, T_{c_1}) P(T_{b_2} < T_{c_2}) + P(T_{a_1} = t_{a_1} < T_{b_1}, T_{c_1}) P(T_{c_2} < T_{b_2}) \\ &= g_{a_1}(t_{a_1})G_{b_1}(t_{a_1})G_{c_1}(t_{a_1}) \left(\int_0^\infty g_{b_2}(t_2)G_{c_2}(t_2) dt_2 + \int_0^\infty g_{c_2}(t_2)G_{b_2}(t_2) dt_2 \right). \end{aligned}$$

This implies that distribution of the completion time of a when it is first in the mimicking parallel system is given by

$$G_{a_1}(t) = \exp \left\{ - \int_{t_{a_1}=0}^t \frac{P(\mathcal{O}_1)f_{a_1}(t_1|\mathcal{O}_1) + P(\mathcal{O}_2)f_{a_1}(t_1|\mathcal{O}_2) + \cdots + P(\mathcal{O}_6)f_{c_1}(t_1|\mathcal{O}_6)}{P(\mathcal{O}_1)F_{a_1}(t_1|\mathcal{O}_1) + P(\mathcal{O}_2)F_{a_1}(t_1|\mathcal{O}_2) + \cdots + P(\mathcal{O}_6)f_{c_1}(t_1|\mathcal{O}_6)} dt_{a_1} \right\}.$$

Hence, while it may seem reasonable to assume that the serial system where if a is first, its completion time has a gamma distribution, if b is first it has a Weibull distribution and if c is first it has a truncated normal distribution, the mimicking parallel system would be required to have a mixture of those distributions for the completion time of a when it finishes first. Nonetheless, there is nothing theoretically preventing a parallel system from having such a distribution.

3.5.1 Distinguishing Parallel and Serial Systems with Selective Influence Manipulations

One limitation of this early methodology was that the mathematical underpinnings that justified the proposed experimental inferences were lacking. Although it had long been known that the expectation of additive random variables, say $T_1 + T_2$, would be an additive function, Sternberg was aware that α_1 might non-selectively affect T_2 , or the other way around. However, it was later shown that even with α_i affecting the proper sub-process directly if, say, T_2 was correlated with T_1 (e.g., T_2 tends to be faster if T_1 was slow and vice versa), then selective influence at the level of the mean RT would usually fail (e.g., Townsend, 1984).

More importantly, if selective influence was assumed to take place at a stronger level, for instance, ordering the distributions such that, say, $F_1(t_1|\alpha_1) > F_1(t_1|\alpha_1^*)$ if and only if $\alpha_1 > \alpha_1^*$, then other architectures could also be tested, in particular parallel architectures (Townsend, 1984, 1990). Furthermore, diverse architectures could be assessed at the far more powerful level of distributional functions (Townsend & Nozawa, 1995; Schweickert, Giorgini, & Dzhafarov, 2000; Dzhafarov *et al.*, 2004).

An important issue has long been how and when selective influence, in the sense of invariance of the marginal distribution of a random completion time (e.g., T_2) marginalized over the other random times (e.g., T_1) might be realized (e.g., Townsend & Ashby, 1983, chapter 11) if sheer stochastic independence was not in force. Although a completely global answer to this question has not yet appeared, Dzhafarov (2003) proposed that the definition of selective influence be as follows:

Definition 3.4 The factors $\alpha_1, \alpha_2, \dots, \alpha_n$ selectively influence the random variables X_1, X_2, \dots, X_n if there exists a set of independent random variables C, S_1, S_2, \dots, S_n that do not change as a function of any α_i , and measurable functions $\lambda_1, \lambda_2, \dots, \lambda_n$ such that

$$X_1 = \lambda_1(\alpha_1, S_1, C), X_2 = \lambda_2(\alpha_2, S_2, C), \dots, X_1 = \lambda_n(\alpha_n, S_n, C).$$

The reader may observe that although the X_1 and X_2 in the definition are not independent, they are conditionally independent (given C). A key consequence

of this conditional independence is that the major theorems on testing of various architectures (Townsend & Nozawa, 1995; Schweickert *et al.*, 2000; Houpt & Townsend, 2011) go through unimpeded (Dzhafarov *et al.*, 2004).

We first consider the connection between selective influence on the object processing times and the order in which those objects are completed. Above, we maintained the possibility that the order in which the processes complete affects completion time distributions in both serial and parallel systems (**SP2**). If a system has **SP2**, then a factor that influences the likelihood of an object finishing in a different order relative to other objects cannot *selectively* influence that completion time. Suppose that the factor associated with X_2 makes it faster so that for all t s $P(T_2^{\text{fast}} \leq t) \geq P(T_2^{\text{slow}} \leq t)$ with strict ordering for at least some t , where $T_2^{\text{fast}} = \lambda_2(\alpha_2^{\text{fast}}, S_2, C)$, $T_2^{\text{slow}} = \lambda_2(\alpha_2^{\text{slow}}, S_2, C)$. Then, in a parallel system, there is a higher chance that object 2 will be faster than object 1, $P(T_2^{\text{fast}} < T_1) > P(T_2^{\text{slow}} < T_1)$, so α_2 can in general influence what position object 1 ends in. If the position in which object 1 finishes affects its completion time distribution, then T_1 with α_2^{fast} must be different from T_1 with α_2^{slow} , but this indicates a failure of selective influence. Alternatively, consider a serial system in which the probability of object 1 being processed first is selectively influenced by α_1 . If the probability of object 1 being first is higher, then the probability of the other objects being first is lower. This means that α_1 affects the position of the other objects, so again, if the completion time distribution of the other objects depends on their positions, then selective influence fails.

With selective influence, we can derive a necessary condition for parallel systems that is not necessary for serial systems based solely on observing the order in which items complete. Selective influence implies marginal selectivity, so the distribution of any pair of completion times (T_i, T_j) does not vary based on factors that selectively influence objects other than i and j . In particular, orderings in a parallel system, i.e., $P(T_i < T_j)$, may be influenced by α_i and α_j but not by other factors. This implies that the probability that object 1 is completed before object 2 is the same regardless of the factors that influence the processing of object 3. Thus, in a parallel system with selective influence, $P(\mathcal{O}(1, 2, 3) \cup \mathcal{O}(1, 3, 2) \cup \mathcal{O}(3, 1, 2))$ is invariant across levels of the factor associated with object 3. In contrast, in a serial process, selective influence on the completion times does not preclude the possibility that the factors influence the processing order (as long as the completion time distributions do not vary based on processing order).

We can derive another test for parallel processing using the total completion times. If the system is parallel, then the total completion time for object i is T_i , which is unaffected by the factor settings of all other objects. If the system is serial, then the total completion time for object i is the sum of all completion times for objects that complete before i and T_i . Suppose α_j influences the processing time of object j ($j \neq i$). If there is some positive probability that object j completes before object i , then the total completion time for object i will sometimes include T_j and hence be influenced by α_j .

Another thing to note about the mapping between parallel and serial systems is that the distribution over orderings in a serial system can be independent of processing time distributions and can even be unrelated to the objects to be processed. On the other hand, in a parallel system, the ordering is dependent on relatively how fast each process completes, which can certainly vary based on object properties (**SP1**, **P3**) and even on other processes (**P2**). Hence, if the order in a serial system does not depend on object properties, then it can be distinguished by examining whether the distribution of order is influenced by changes in the object properties. This leads us into the next section on distinguishing serial and parallel systems when there is an option to selectively influence a process.

Example 3.4 As a special case, there is the situation when the processing order is deterministic, e.g., a particular order has probability 1. Intuitively, this should only be possible for serial systems. However, the outlined properties do not trivially constrain parallel processes so much that they cannot have this property. Consider a three sub-process system as in Example 3.3. Thus, the stopping regions are [1, 2], [3, 4], and [5, 6] for each respective sub-process X_1 , X_2 , and X_3 . Now allow the sub-processes to have the following construction:

- $X_1(t)$ follows a *unif*(0, 1) distribution for each $t \leq 2$ and follows *unif*[1, 2] distribution for $t > 2$.
- $X_2(t)$ follows a *unif*(0, 1) distribution for each $t \leq 3$ and follows *unif*[3, 4] distribution for $t > 3$.
- $X_3(t)$ follows a *unif*(0, 1) distribution for each $t \leq 4$ and follows *unif*[5, 6] distribution for $t > 4$.

Each X_i is being worked on from time 0 until it is complete. Also, each sub-process X_i is guaranteed (probability = 1) to finish at time $t = i + 1$. Thus, the processing order is deterministic. Note that modifying the definition of “being worked on” to require that the process could finish at any time, not just change state, would negate this example.

Finally we come to the classic survivor interaction contrast tests developed by Townsend and Nozawa (1995) and extended by Schweickert, Dzhafarov and colleagues (Schweickert *et al.*, 2000, 2012; Dzhafarov *et al.*, 2004; Zhang & Dzhafarov, 2015).

Consider first a system with two independent sub-processes that can be sped up and slowed down by selective influence manipulations. The survivor interaction contrast is defined by an interaction contrast of the survivor functions of the system completion times under two levels of the selective influence manipulation on each sub-process. In particular, suppose there is some manipulation that speeds up or slows down each process selectively,

$$\begin{aligned} T_1^H &= \lambda_1(\alpha_1^H, S_1) \\ T_1^L &= \lambda_1(\alpha_1^L, S_1) \end{aligned}$$

such that for all t , $P(T_1^H > t) \leq P(T_1^L > t)$ with strict inequality for some t , i.e., T_1^H stochastically dominates T_1^L .

The survivor interaction contrast is given by

$$\text{SIC}(t) = (S^{LL} - S^{LH}) - (S^{HL} - S^{HH}) = (F^{HL} - F^{HH}) - (F^{LL} - F^{LH}).$$

In a parallel system that stops as soon as either process has finished and independence between T_1 and T_2 , the survivor function for the system completion time is given by the probability that neither of the processes have completed by time t ,

$$S(t) = S_1(t)S_2(t).$$

Hence, the survivor interaction contrast can be rewritten as

$$\begin{aligned} \text{SIC}(t) &= (S_1^L(t)S_2^L(t) - S_1^L(t)S_2^H(t)) - (S_1^H(t)S_2^L(t) - S_1^H(t)S_2^H(t)) \\ &= S_1^L(t)(S_2^L(t) - S_2^H(t)) - S_1^H(t)(S_2^L(t) - S_2^H(t)) \\ &= (S_1^L(t) - S_1^H(t))(S_2^L(t) - S_2^H(t)). \end{aligned}$$

Because of the stochastic dominance assumption, each term in the product must be positive so the product is positive and hence the survivor interaction contrast for a parallel, first-terminating process is positive for all t . Intuitively, we might expect this from the interaction contrast because the minimum of the two sub-processes should be affected more by going from L to H when the other sub-process is L compared to going from L to H when the other sub-process is already H .

If the parallel system stops only when both sub-processes complete, again assuming independence between T_1 and T_2 , then the survivor function for the system completion time is given by the probability that neither sub-process is still continuing,

$$S(t) = 1 - (1 - S_1(t))(1 - S_2(t)) = S_1(t) + S_2(t) - S_1(t)S_2(t).$$

The survivor interaction contrast is given by

$$\begin{aligned} \text{SIC}(t) &= (S_1^L(t) + S_2^L(t) - S_1^L(t)S_2^L(t) - (S_1^L(t) + S_2^H(t) - S_1^L(t)S_2^H(t))) \\ &\quad - (S_1^H(t) + S_2^L(t) - S_1^H(t)S_2^L(t) - (S_1^H(t) + S_2^H(t) - S_1^H(t)S_2^H(t))) \\ &= (-S_1^L(t)S_2^L(t) + S_1^L(t)S_2^H(t)) - (-S_1^H(t)S_2^L(t) + S_1^H(t)S_2^H(t)) \\ &= S_1^L(t)(-S_2^L(t) + S_2^H(t)) - S_1^H(t)(-S_2^L(t) + S_2^H(t)) \\ &= (S_1^L(t) - S_1^H(t))(S_2^H(t) - S_2^L(t)). \end{aligned}$$

Again, relying on stochastic dominance from the selective influence manipulation, the first term in the final product is positive for all t while the second term is negative for all t , indicating that the survivor interaction contrast for the parallel, exhaustive condition is negative for all t . The intuition for this result mirrors the intuition for the parallel, first-terminating model. If the slowest of the processes determines the system completion time, then going from L to H when the other process is already L will have less of an effect than going from L to H when the other process is H .

In a serial system that stops as soon as either sub-process is finished that has some probability p of a particular sub-process being processed first (and assuming that the selective influence manipulations on the sub-process completion times do not affect p), the completion time distribution is

$$F(t) = pF_1(t) + (1 - p)F_2(t).$$

Hence, the survivor interaction contrast is

$$\begin{aligned} \text{SIC}(t) &= (pF_1^H(t) + (1 - p)F_2^L(t) - pF_1^H(t) - (1 - p)F_2^H(t)) \\ &\quad - (pF_1^L(t) + (1 - p)F_2^L(t) - pF_1^L(t) - (1 - p)F_2^H(t)) \\ &= ((1 - p)F_2^L(t) - (1 - p)F_2^H(t)) - ((1 - p)F_2^L(t) \\ &\quad - (1 - p)F_2^H(t)) = 0. \end{aligned}$$

Finally, for the serial model which requires both sub-processes to finish, the distribution is the convolution of the completion time density of one sub-process with the completion time distribution of the other sub-process,

$$S(t) = \int_0^\infty f_1(s)S_2(t-s) ds = f_1 * S_2.$$

This means the survivor interaction contrast is given by

$$\begin{aligned} \text{SIC}(t) &= \left(\int_0^\infty f_1^L(s)S_2^L(t-s) ds - \int_0^\infty f_1^L(s)S_2^H(t-s) ds \right) \\ &\quad - \left(\int_0^\infty f_1^H(s)S_2^L(t-s) ds - \int_0^\infty f_1^H(s)S_2^H(t-s) ds \right) \\ &= \int_0^\infty (f_1^L(s)(S_2^L(t-s) - S_2^H(t-s)) \\ &\quad - f_1^H(s)(S_2^L(t-s) - S_2^H(t-s))) ds \\ &= \int_0^\infty (f_1^L(s) - f_1^H(s))(S_2^L(t-s) - S_2^H(t-s)) ds. \end{aligned}$$

Due to the stochastic dominance from the selective influence manipulation, the difference of survivor functions in the product will always be positive, so the difference of densities determines the sign of the integrand. Because of the dominance on the first sub-process, $f_1^L(t)$ must be smaller than $f_1^H(t)$ for some range of time $(0, t^*)$, so the integrand, and hence the SIC, is negative for some initial range of times. To verify that the SIC is also positive for some range of time, it is sufficient to show that the integrated SIC is 0. Because of the relationship (for positive random variables), $\int_0^\infty S_X(t) dt = E[X]$, the integrated survivor interaction contrast is the interaction contrast of the means,

$$\int_0^\infty \text{SIC}(t) dt = (E[T^{\text{LL}}] - E[T^{\text{LH}}]) - (E[T^{\text{HL}}] - E[T^{\text{HH}}]).$$

Because the expectation is a linear operator, $E[T] = E[T_1] + E[T_2]$, so

$$\begin{aligned} \int_0^\infty \text{SIC}(t) dt &= (E[T_1^L] + E[T_2^L] - E[T_1^L] - E[T_1^H]) \\ &\quad - (E[T_1^H] + E[T_2^L] - E[T_1^H] - E[T_1^H]) = 0. \end{aligned}$$

In more recent work, Yang, Fifić, and Townsend (2012) demonstrated that, for most cases, there is only a single zero-crossing for the serial-exhaustive survivor interaction contrast. For our current purposes, it is sufficient that we have demonstrated that using the survivor interaction contrast with selective influence manipulations, serial and parallel processes can be distinguished.

Although these proofs all depended on independence between the completion time distributions of the sub-processes, as long as selective influence holds, the same conclusions hold. By the selective influence definition above, the completion time of one sub-process is *conditionally* independent of the other sub-process completion time given C . Hence, if we replace the densities, distributions and survivor functions above with the conditional versions, then the derivations hold for the dependent, but selectively influenced, processes. To extend from there to the unconditional distributions, note that the conditional holds when conditioned on any C so the derivations above are true when integrated across all C , i.e., when C is marginalized.

3.6 Discussion and Conclusions

In this chapter, it has been our goal to formalize the concepts of parallel and serial systems in cognitive modeling. Based on this foundation, we have endeavored to enumerate the contexts in which these systems are discriminable and those contexts in which they can perfectly mimic one another.

This chapter is by no means the first treatise on this topic. Studies concerning the identifiability of serial and parallel systems have appeared since the early days of cognitive psychology. We have attempted to draw attention to the landmarks in the development of this important topic, but because our goal was to illuminate the issues of serial and parallel identifiability themselves rather than their history, we have necessarily omitted direct reference to important works in this domain.

What we have done is to begin with a modification of the informal statement of the shared and distinctive properties of parallel and serial systems posited by Townsend and Ashby (1983). From there, we gave a brief overview of the necessary measure-theoretic probability theory to introduce the concepts of filtrations and adapted stochastic processes. We then used the formalism of an adapted stochastic process as the foundation of our definitions of parallel and serial systems. The use of adapted stochastic processes has allowed us to formalize the

notion that these systems should maintain causality, particularly the constraint that the future should not influence the present. These formal systems were then used to explore the necessary (and sometimes sufficient) conditions under which serial and parallel systems may be discriminated. In many cases, serial and parallel structures are only discriminable under selective influence manipulations, including some of the most-used theorems that we covered, those concerning the survivor interaction contrast.

As part of our development, we have also highlighted many instances in which parallel and serial systems are *not* discriminable. This may serve as a cautionary note to empirical researchers not to quickly label cognitive systems as parallel or serial without considering these possible mimicking issues.

On a more positive note, we hope that these theoretical developments may serve as a foundation for future research that might result in more, and more powerful, tests for empirically discriminating between these two classes of systems.

References

- Algom, D., Eidels, A., Hawkins, R. X. D., Jefferson, B., & Townsend, J. T. (2015). Features of response times: Identification of cognitive mechanisms through mathematical modeling. In *Oxford library of psychology. The Oxford handbook of computational and mathematical psychology* (pp. 63–98). Oxford: Oxford University Press.
- Billingsley, P. (1995). *Probability and measure*, third edition. New York, NY: Wiley.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178.
- Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika*, 68, 7–26.
- Dzhafarov, E. N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, 50, 390–401.
- Dzhafarov, E. N., & Kujala, J. V. (2010). The joint distribution criterion and the distance tests for selective probabilistic causality. *Frontiers in Psychology*, 1, 151.
- Dzhafarov, E. N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, 48, 51–64.
- Houpt, J. W., & Townsend, J. T. (2011). An extension of SIC predictions to the Wiener coactive model. *Journal of Mathematical Psychology*, 55, 267–270.
- Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology*, 52, 128–144.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108.
- Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a stroop task. *Journal of Mathematical Psychology*, 18, 105–139.
- Schweickert, R., Fisher, D. L., & Sung, K. (2012). *Discovering cognitive architecture by selectively influencing mental processes* (Vol. 4). Singapore: World Scientific.

- Schweickert, R., Giorgini, M., & Dzhafarov, E. (2000). Selective influence and response time cumulative distribution functions in serial–parallel task networks. *Journal of Mathematical Psychology*, 44, 504–535.
- Snodgrass, J. G., & Townsend, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 330–354.
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 4, 421–457.
- Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (ed.) *Attention and performance*, vol. 4 (pp. 629–685). New York: Academic Press.
- Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25, 168–199.
- Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–168). Hillsdale, NJ: Erlbaum Press.
- Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. *Journal of Mathematical Psychology*, 14, 219–238.
- Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, 28, 363–400.
- Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedle-dum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, 1, 46–54.
- Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- Townsend, J. T., & Evans, R. (1983). A systems approach to parallel-serial testability and visual feature processing. In H. G. Geissler (ed.), *Modern issues in perception* (pp. 166–189). Berlin: VEB Deutscher Verlag der Wissenschaften.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, 39, 321–360.
- Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, 33, 309–327.
- Townsend, J. T., & Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology*, 38, 1–24.
- Van Zandt, T., & Townsend, J. T. (2012). Designs for and analyses of response time experiments. In T. K. Little (ed.), *The Oxford handbook of quantitative methods* (pp. 260–285). New York, NY: Oxford University Press.
- Yang, H., Fifić, M., & Townsend, J. T. (2012). Survivor interaction contrast wiggle predictions of parallel and serial models for an arbitrary number of subprocesses. *Journal of Mathematical Psychology*, 48, 21–32.
- Zhang, R., & Dzhafarov, E. (2015). Noncontextuality with marginal selectivity in reconstructing mental architectures. *Frontiers in Psychology*, 6(735). Available from www.frontiersin.org/cognition/10.3389/fpsyg.2015.00735/abstract

4 Identifiability of Probabilistic Models, with Examples from Knowledge Structure Theory

Jean-Paul Doignon, Jürgen Heller, and Luca Stefanutti

4.1	An Example and Our Goals in the Chapter	129
4.2	Probabilistic Models	132
4.3	Mathematical Tools	140
4.3.1	Convex Polytopes	140
4.3.2	Two Theorems (and More Propositions) from Mathematical Analysis	141
4.3.3	Investigating Probabilistic Models	145
4.3.4	Results about Model Testability	146
4.3.5	Results about Model Identifiability	146
4.3.6	Summary	147
4.4	Knowledge Structures: Combinatorial Aspects	148
4.5	Probabilistic Knowledge Structures	152
4.6	The Correct Response Model: Testability	153
4.7	The Correct Response Model: Identifiability	158
4.8	The Basic Local Independence Model	162
4.8.1	Parameter Domain	163
4.8.2	Outcome Space	164
4.8.3	Prediction Function	164
4.9	The Basic Local Independence Model: Testability	165
4.10	The Basic Local Independence Model: Identifiability	166
4.10.1	Trade-off Dimensions	173
4.11	Latent Class Models	176
4.12	Conclusions	180
4.13	Bibliographical Notes	181
	Acknowledgments	182
	References	182

4.1 An Example and Our Goals in the Chapter

Many models in mathematical psychology aim at explaining the frequency of observed behavior. They define the probabilities of occurrence of those behaviors on the basis of latent attributes and mechanisms. Fundamental questions about such probabilistic models raise interesting, mathematical problems. Let us take an elementary toy example to provide an illustration and give a first hint at the themes of the chapter.

Suppose students repeatedly take a test, consisting of only two items bearing each one on some notion in arithmetics. For instance, the two notions could be

a: “addition of two positive integers less than 10”

and

b: “multiplication of two positive integers less than 99.”

To repeatedly test the mastery of a notion, an item formulates the same problem in various disguises. In the case of the addition item *a* above, the problem submitted to the student could be “8 + 7,” or “2 + 6,” or “9 + 9,” etc. For any student, we record the relative frequency of her correct responses on each item *q*, that is, the number of times the answer she provides is correct, divided by the total number of her responses. In our toy example, the data could consist of the pair (f_a, f_b) of observed, relative frequencies of correct answers by the student to the two items *a* and *b* (later in this section we consider another, related type of data for the same toy example).

There are models explaining the occurrence of such frequencies f_a and f_b (their general formulation is not restricted to the case of two items). One of these models relies on the following natural assumptions. At each time she provides an answer, the student knows (masters) certain of the notions. In other words, her knowledge state is described as some subset of the whole set of notions covered by the test (here, a subset of $\{a, b\}$); we call this whole set the (knowledge) domain. It is mainly the actual knowledge state of the student that governs the correctness of any answer from the student – of course, we need to specify how, and we will explain this in a few lines. We consider that the student’s knowledge state may be drawn from a fixed collection of subsets of the domain, according to some probability distribution on the collection. The probability distribution belongs to the latent part of the model: indeed, the only observables are the frequencies of correct responses to the items by a given student (or in the second model we will consider, the frequencies of response patterns).

The above elementary assumptions lead to the following formalization for our toy example. The domain is $Q = \{a, b\}$. All possible knowledge states form a certain collection \mathcal{K} of subsets of the domain. To make room for a completely ignorant student, we make sure that the empty set belongs to \mathcal{K} ; similarly, we assume $Q \in \mathcal{K}$ for a student knowing all of the items. To take an example here, suppose there is only one further knowledge state, namely $\{a\}$, so that

$$\mathcal{K} = \{\emptyset, \{a\}, Q\}. \quad (4.1)$$

Next, we assume that some probability distribution π exists on the collection \mathcal{K} , with π_K the probability of the state K . That is, for any K in \mathcal{K} , the real number π_K satisfies

$$0 \leq \pi_K \leq 1 \quad (4.2)$$

with moreover

$$\pi_\emptyset + \pi_{\{a\}} + \pi_Q = 1. \quad (4.3)$$

The correctness of a student's answer to some item depends on her knowledge state, but there could be other influences to take into account. First, even if the item is in principle mastered, the student might provide a wrong answer because of absent-mindedness. Second, even if the item is outside her knowledge state, she could provide the correct answer just by a lucky guess (this is especially true if the item proposes a small number of possible answers, as in a multiple choice test). The model thus introduces for each item q two numbers β_q and η_q , the probability of a careless error and the probability of a correct guess on that item. Then it expresses the probability of a correct answer by conditioning on the actual state of the student; when the state is selected, the probability of a correct response follows from the interpretations of the error and guessing probabilities β_q and η_q . Assuming various independence assumptions, we derive the equation giving the probability $\tau(q)$ of a correct response to item q . In our example, we only have two items a and b , and we set:

$$\tau(a) = \eta_a \pi_\emptyset + (1 - \beta_a) \pi_{\{a\}} + (1 - \beta_a) \pi_Q, \quad (4.4)$$

$$\tau(b) = \eta_b \pi_\emptyset + \eta_b \pi_{\{a\}} + (1 - \beta_b) \pi_Q. \quad (4.5)$$

The probabilistic model we obtain for (the probabilities of) correct answers is designated as the correct response model (CRM), and denoted as CRM(\mathcal{K}).

Still working with the above toy example of a probabilistic model, we introduce a few basic concepts. Notice first that the model has $3 + 4$ parameters π_\emptyset , $\pi_{\{a\}}$, π_Q , η_a , η_b , β_a , β_b . In view of their interpretations, each parameter may take values only in a restricted set of real numbers, for instance, $0 \leq \eta_a \leq 1$ and similarly for η_b , β_a , β_b ; moreover, we keep Equations (4.2) and (4.3). All the allowed 7-tuples of real numbers form the so-called "parameter domain." Formulas (4.4) and (4.5) are at the core of the model: for each 7-tuple in the parameter domain, they define the corresponding set of predicted values for the observables $\tau(a)$, $\tau(b)$. The resulting 2-tuples $(\tau(a), \tau(b))$ form the "prediction range" of the model (notice that $0 \leq \tau(a) \leq 1$ and $0 \leq \tau(b) \leq 1$ follow from Equations (4.4) and (4.5)). We now check whether our toy model satisfies some general properties.

We first rewrite Formulas (4.4) and (4.5) in the case where π is concentrated on the empty state (that is, $\pi_\emptyset = 1$):

$$\tau(a) = \eta_a, \quad (4.6)$$

$$\tau(b) = \eta_b. \quad (4.7)$$

If the values of the guessing probabilities η_a and η_b are not restricted in any way, we thus see that the model produces all real values from 0 to 1 for each of the two correct response probabilities $\tau(a)$ and $\tau(b)$. Hence, it explains all conceivable data, with the consequence that it can never be refuted. Such an “untestable” model is only of limited interest. If we allow $\pi_\emptyset = 1$ in the model, we thus need to set some restrictions on the values of the parameters η_a and η_b (this makes sense also in view of the interpretation of η_a and η_b , and will be done later on). From the case $\pi_Q = 1$, a similar conclusion is reached for the parameters β_a and β_b .

The model has still another particular feature: the same potential data (or predicted point) $(\tau(a), \tau(b))$ can be produced by several 7-tuples of allowed values for the seven parameters $\pi_\emptyset, \pi_{\{a\}}, \pi_Q, \beta_a, \beta_b, \eta_a, \eta_b$. This feature might be guessed from the fact that our data consist of only two real values, while there are as many as six non-redundant real parameters (in this respect, see Corollary 4.24). To give a formal argument, let us rewrite the equations in the case that $\pi_\emptyset = \pi_{\{a\}} = \pi_Q = 1/3$:

$$\tau(a) = (\eta_a + 2 - 2\beta_a)/3, \quad (4.8)$$

$$\tau(b) = (2\eta_b + 1 - \beta_b)/3. \quad (4.9)$$

We see that it suffices to keep each of $\eta_a - 2\beta_a$ and $2\eta_b - \beta_b$ constant to get the same values for $\tau(a)$ and $\tau(b)$. The model thus explains the same data with many different vectors¹ of parameter values; we say that it is “non-identifiable.” Some researchers deem that non-identifiability makes the model without any value, while others do not worry about this. Anyway, it is often important to decide for a given probabilistic model whether it is identifiable or not.²

Here is another model, which differs from the CRM(\mathcal{K}) above in that it predicts patterns of correct responses rather than correct responses separately. We consider again the domain $Q = \{a, b\}$, the collection $\mathcal{K} = \{\emptyset, \{a\}, Q\}$ of knowledge states and the same seven parameters as for CRM(\mathcal{K}). A ‘response pattern’ of a student is any subset of Q , which we interpret as consisting of the items on which (at the considered moment) this student would provide correct answers. For a two-item domain, there are thus four patterns of responses. Implementing the same natural assumptions as in the CRM, we come to the following predicted probabilities for the four patterns:³

$$\rho(\emptyset) = (1 - \eta_a)(1 - \eta_b)\pi_\emptyset + \beta_a(1 - \eta_b)\pi_{\{a\}} + \beta_a\beta_b\pi_Q, \quad (4.10)$$

$$\rho(\{a\}) = \eta_a(1 - \eta_b)\pi_\emptyset + (1 - \beta_a)(1 - \eta_b)\pi_{\{a\}} + (1 - \beta_a)\beta_b\pi_Q, \quad (4.11)$$

$$\rho(\{b\}) = (1 - \eta_a)\eta_b\pi_\emptyset + \beta_a\eta_b\pi_{\{a\}} + \beta_a(1 - \beta_b)\pi_Q, \quad (4.12)$$

$$\rho(Q) = \eta_a\eta_b\pi_\emptyset + (1 - \beta_a)\eta_b\pi_{\{a\}} + (1 - \beta_a)(1 - \beta_b)\pi_Q. \quad (4.13)$$

¹ The elements of a linear space \mathbb{R}^m are denominated indifferently as points or vectors.

² Particularly if statistical investigations are made on the model, like searching for the maximum likelihood estimates of its parameters.

³ It is an exercise to check that ρ as given by Equations (4.10)–(4.13) is indeed a probability distribution on the set of four response patterns.

This is a particular basic local independence model for response patterns (BLIM), denoted as $\text{BLIM}(\mathcal{K})$, a probabilistic model that we will later investigate in depth as regards testability and identifiability (see Sections 4.9 and 4.10).

Assessing the actual values of the parameters in a realization of $\text{BLIM}(\mathcal{K})$ requires that data provide the frequencies for all of the $2^{|Q|}$ response patterns (a huge number if $|Q|$ is not small, and moreover we need to evaluate the correctness of the responses of a student in a short period of time to avoid changes in her state).

To get reliable estimates of the frequencies of the response patterns, a sufficiently large sample of students is needed. The task is thus feasible only for small numbers of items. Of course, computerized testing helps here; with thousands of students, one can handle a dozen of items. Notice that collecting data for $\text{CRM}(\mathcal{K})$ is easier, because we need only collect $|Q|$ frequencies of correct responses to the various items, and to this aim we may handle items separately.

In the chapter we state formal definitions first for a “probabilistic model,” and then for several interesting properties that such a model can have. We thus introduce in a precise way the notions of “testability” and “identifiability” as well as a few of their variants. We also describe some of the mathematical tools available for testing whether a model has such a property. After having defined “probabilistic knowledge structures,” we make use of them not only to provide illustrations of the concepts and tools, but also to state results and some open problems about identifiability. As a matter of fact, we concentrate on two models for student answers that emerged from knowledge structure theory (our toy example gave two particular instances of them): the correct response model (CRM) and the basic local independence model (BLIM).

Here is how the chapter is organized. Section 4.2 defines probabilistic models, as well as testability and identifiability (together with several variants). Section 4.3 surveys some relevant mathematical tools. Sections 4.4 and 4.5 form a short introduction to knowledge structure theory (KST). Results about the CRM appear in Sections 4.6 and 4.7. Those about the BLIM follow in Sections 4.8 to 4.10. The next Section 4.11 mentions latent class models and explains their relationship with the BLIM. After a conclusion in Section 4.12, our final Section 4.13 provides bibliographic orientation; the list of references follows.

4.2 Probabilistic Models

Probabilistic models serve to predict some behavior or, more precisely, the stochastic realization of a collection of observable variables. They express the probability of any combination of values of the observables⁴ as a function of the parameter values (that is, of a parameter point); the latter captures some unknown, latent information (which can be of various forms). The term ‘predicted point’ designates “any set of probability values for the observable outcomes that

⁴ Or more generally of any “measurable” set of observable values. However, in the chapter, we only consider finite sets of behavior values, and may thus ignore the technical points of measure theory.

the model predicts.” Here are two fundamental questions about a probabilistic model:

- (i) MODEL TESTABILITY: is there some distribution of probability values for the observables that the model does not predict?
- (ii) MODEL IDENTIFIABILITY: is each predicted distribution produced from at most one parameter point?

The aim of this section is to state precise definitions of the notions we just evoked. We mainly follow the expositions in Bamber and van Santen (1985, 2000), although our first three definitions (Definitions 4.1–4.3) are stated in a more general context.

Definition 4.1 A model $(\mathcal{D}, f, \mathcal{O})$ is a triple consisting of a *parameter domain* \mathcal{D} , an *outcome space* \mathcal{O} , and a *prediction function*

$$f : \mathcal{D} \rightarrow \mathcal{O}, \quad (4.14)$$

with $f(\mathcal{D})$ the *prediction range*. We call *parameter points* the elements of \mathcal{D} , *outcome points* those of \mathcal{O} and *predicted points* those of $f(\mathcal{D})$. The model is *probabilistic* when the outcome points are probability distributions on the same finite set of elementary events,⁵ or families of such probability distributions.

In the toy example of Section 4.1, the outcome points of CRM(\mathcal{K}) consist of a pair of probability distributions, each one on two elementary outcomes: a correct response and a false response to the (first, vs. second) item; however, we usually mention only the probability of the correct answer (because the probability of a false answer follows at once). The outcomes of BLIM(\mathcal{K}) are probability distributions on the collection of the four response patterns $\emptyset, \{a\}, \{b\}, Q$. For any model a predicted point is always an outcome point, but the converse does not necessarily hold (see Example 4.7 below). If some outcome point not predicted by the model appears as data, we would conclude that the model does not explain all data points.

Definition 4.2 The model $(\mathcal{D}, f, \mathcal{O})$ is *testable* if $f(\mathcal{D}) \subset \mathcal{O}$ (strict inclusion).

We now turn to the concept of an identifiable model.

Definition 4.3 The model $(\mathcal{D}, f, \mathcal{O})$ is *identifiable* if f is injective (that is, one-to-one from \mathcal{D} to $f(\mathcal{D})$). It is *identifiable at the parameter point* x in \mathcal{D} if for any x' in \mathcal{D} , the equality $f(x) = f(x')$ implies $x = x'$. Then we also say that the model is *identifiable at the predicted point* $f(x)$ in \mathcal{O} .

Notice that a model is identifiable if and only if it is identifiable at every point of its parameter domain. However, a non-identifiable model may well be identifiable at some points of its parameter domain (see Example 4.7).

⁵ In a more general setting, there may be infinitely many elementary events, and then probability distributions must be replaced with more sophisticated tools from probability theory, such as probability densities.

Many researchers consider that a model, to be interesting, must be identifiable. They reject the strange feature at the core of non-identifiability: the same real situation, or predicted point, is explained by distinct parameter points. Let us now explain how any non-identifiable model $(\mathcal{D}, f, \mathcal{O})$ can always be transformed into an identifiable model having the same prediction range. It suffices to restrict the parameter domain in a specific way. Indeed, take the new parameter domain \mathcal{E} as any subset of \mathcal{D} containing exactly one element per preimage $f^{-1}(b)$, for $b \in f(\mathcal{D})$. Then define the new prediction function g as the restriction of f to \mathcal{E} . The resulting model $(\mathcal{E}, g, \mathcal{O})$ is identifiable (indeed, g is injective), with moreover $g(\mathcal{E}) = f(\mathcal{D})$. There are generally many possible choices for \mathcal{E} . In practical situations, it is not always obvious how to select the subset \mathcal{E} in a meaningful way. We will discuss this question for the CRM in Section 4.7.

We now indicate how the CRM and the BLIM based on the toy model from Section 4.1 fall under Definition 4.1. We also check whether they are testable and identifiable.

Example 4.4 In the CRM based on the toy example of Section 4.1, the parameter domain is

$$\begin{aligned} \mathcal{D} = \{(\pi_{\emptyset}, \pi_{\{a\}}, \pi_Q, \eta_a, \eta_b, \beta_a, \beta_b) \in [0, 1]^7 \mid \\ \pi_{\emptyset} + \pi_{\{a\}} + \pi_Q = 1\}. \end{aligned} \quad (4.15)$$

The outcome points are pairs of real numbers in $[0, 1]$, each one capturing the probability of a correct answer to an item (either item a or item b). Hence we consider that the outcome space is the real unit square $\mathcal{O} = [0, 1] \times [0, 1]$. The prediction function is the mapping

$$f : \mathcal{D} \rightarrow \mathcal{O} : (\pi_{\emptyset}, \pi_{\{a\}}, \pi_Q, \eta_a, \eta_b, \beta_a, \beta_b) \mapsto (\tau(a), \tau(b)) \quad (4.16)$$

where the pair $(\tau(a), \tau(b))$ is given by Equations (4.4) and (4.5). As we saw in Section 4.1, it happens in this model CRM(\mathcal{K}) that all outcome points are also predicted points. In other words, this model (based on two items) is not testable. As seen also in Section 4.1, the same model is not identifiable. We now prove a stronger assertion: this model is identifiable at none of its parameter points. It suffices to prove that it is not identifiable at any of its predicted points, say $(\tau(a), \tau(b))$. The latter is the image by the prediction function f of any parameter point satisfying $\pi_{\emptyset} = 1$, $\eta_a = \tau(a)$ and $\eta_b = \tau(b)$, with any value of β_a and β_b .

Example 4.5 The BLIM(\mathcal{K}) based on the toy example of Section 4.1 has the same parameter domain as the CRM(\mathcal{K}) in Example 4.4. Its outcome points are quadruples of real numbers in $[0, 1]$ indexed by the four response patterns and summing up to 1: each quadruple stores one of the probability distributions on the set of four response patterns, so that $\mathcal{O} = \{p \in [0, 1]^4 \mid \sum p = 1\}$. The prediction function is the mapping

$$f : \left\{ \begin{array}{l} \mathcal{D} \rightarrow \mathcal{O} : \\ (\pi_{\emptyset}, \pi_{\{a\}}, \pi_Q, \eta_a, \eta_b, \beta_a, \beta_b) \mapsto (\rho(\emptyset), \rho(\{a\}), \rho(\{b\}), \rho(Q)) \end{array} \right. \quad (4.17)$$

where the values of ρ are given by Equations (4.10)–(4.13). Is this model testable? The answer is negative, as we now show by proving that any point p in \mathcal{O} equals some $(\rho(\emptyset), \rho(\{a\}), \rho(\{b\}), \rho(Q))$ for an adequate choice of the parameter point. If we set $\eta_a = 0$ and $\beta_a = 0$, the other parameters still need to satisfy

$$\begin{cases} p(\emptyset) = (1 - \eta_b)\pi_{\emptyset} \\ p(\{a\}) = (1 - \eta_b)\pi_{\{a\}} + \beta_b\pi_Q \\ p(\{b\}) = \eta_b\pi_{\emptyset} \\ p(Q) = \eta_b\pi_{\{a\}} + (1 - \beta_b)\pi_Q. \end{cases} \quad (4.18)$$

We may choose values for the other parameters as follows. First, we set $\pi_{\emptyset} = p(\emptyset) + p(\{b\})$, next $\eta_b = \frac{p(\{b\})}{p(\emptyset) + p(\{b\})}$ (or $\eta_b = 0$ if $p(\emptyset) + p(\{b\}) = 0$). Second, we set $\pi_{\{a\}} = 0$ and $\pi_Q = p(\{a\}) + p(Q)$, finally $\beta_b = \frac{p(\{a\})}{p(\{a\}) + p(Q)}$ (or $\beta_b = 0$ if $p(\{a\}) + p(Q) = 0$). It is easily checked that the resulting 7-tuple of values so assigned to $(\pi_{\emptyset}, \pi_{\{a\}}, \pi_Q, \eta_a, \eta_b, \beta_a, \beta_b)$ gives a point in the parameter domain \mathcal{D} of the model. Moreover, the image of this 7-tuple by f is the given outcome point specified by p .

Is the model identifiable? The answer is no, because the same predicted point y can be obtained by letting $\eta_a = 1$, $\beta_a = 1$ and $\pi_{\{a\}} = 0$, and then choosing the values of the four other parameters in order to satisfy

$$\begin{cases} p(\emptyset) = \beta_b\pi_Q \\ p(\{a\}) = (1 - \eta_b)\pi_{\emptyset} \\ p(\{b\}) = (1 - \beta_b)\pi_Q \\ p(Q) = \eta_b\pi_{\emptyset} \end{cases} \quad (4.19)$$

(there is a 7-tuple of values that defines a parameter point of the BLIM while satisfying the above equations). The latter argument indicates that $\text{BLIM}(\mathcal{K})$ (based on the toy example) is identifiable at no predicted point, thus at no parameter point.

Examples 4.4 and 4.5 show that, on our toy example of a knowledge structure with only two items, both the CRM and the BLIM are of limited interest (they are both untestable and non-identifiable). We will later see that, for many other knowledge structures (having a larger number of items in their domains), the conclusions are different.

In many models (as is the case in Examples 4.4 and 4.5) the parameter points lie in some fixed numerical space \mathbb{R}^m , where $m \in \mathbb{N}$; then, their components are called the *parameters*. This occurs for sure when the parameter points are probability distributions on some fixed set of cardinality m . Another case is when the parameter points are the vectors of parameters of some family of distributions (for instance, the vector formed by the mean and the standard deviation of a normal distribution). Similarly the predicted points often lie in \mathbb{R}^n for some n in \mathbb{N} . In the BLIM, each predicted point is a probability distribution on the finite set of response patterns. In the CRM, a component of a predicted point is a probability value summarizing a probability distribution on two elementary outcomes (false and correct responses).

To introduce more concepts about models, we assume first that the outcome space lies in some \mathbb{R}^n , second that moreover the parameter domain lies in some \mathbb{R}^m . The use of the Lebesgue measure of a (measurable) subset of \mathbb{R}^n should not frighten the reader: let us agree that it is just a fancy name for the volume of well-behaved subsets of \mathbb{R}^n (of course, since Lebesgue, a perfectly rigorous definition is available – see any manual on advanced analysis). For a subset X of \mathbb{R}^n , we write $\mu(X)$ to denote the Lebesgue measure of X (under the implicit assumption that this particular subset X is measurable).

Definition 4.6 Assume $\mathcal{O} \subseteq \mathbb{R}^n$, for some n in \mathbb{N} , and let μ be the (Lebesgue) measure in \mathbb{R}^n . The model $(\mathcal{D}, f, \mathcal{O})$ is *quantitatively testable* if the measure of \mathcal{O} is strictly positive, while $f(\mathcal{D})$ has null measure. In formulas: $\mu(\mathcal{O}) > \mu(f(\mathcal{D})) = 0$.

Finding a data point in the prediction range of a model which is quantitatively testable is much more informative than if the model were just testable. Indeed, if the model is quantitatively testable, a data point sampled randomly from the outcome space⁶ almost never falls in the predictive range.

Example 4.7 Assume $m = n = 1$, $\mathcal{D} = [0, 1]$ and $\mathcal{O} = [0, 1]$. Take the prediction function $f : [0, 1] \rightarrow [0, 1] : x \mapsto 2(x - 0.4)^2$ (see Figure 4.1 for an illustration). The resulting model $(\mathcal{D}, f, \mathcal{O})$ is testable, because $f(\mathcal{D}) = [0, 0.72]$ is a strict subset of $\mathcal{O} = [0, 1]$. It is not quantitatively testable, because $\mu(f(\mathcal{D})) = 0.72 > 0$ (here μ is the measure on the real line).

The model is identifiable at any parameter point in $]0.8, 1]$, but at no other point, and so it is not identifiable. To build an identifiable model with the same prediction range, we may define a new parameter domain $\mathcal{E} = [0.4, 1]$ and take g to be the restriction of f to \mathcal{E} . Then the model $(\mathcal{E}, g, \mathcal{O})$ is identifiable and $f(\mathcal{D}) = g(\mathcal{E}) = [0, 0.72]$. We could rather set $\mathcal{E} = [0, 0.4] \cup]0.8, 1]$, and again

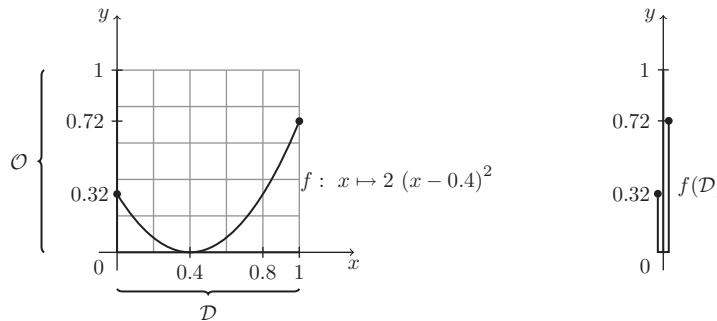


Figure 4.1 (Left) The graph of the prediction function in Example 4.7. (Right) A pictorial representation of the prediction range (with marked endpoints) in the outcome space ($[0, 1]$).

⁶ w.r.t. a probability measure on the outcome space having a density function which is positive for all points.

restrict f to \mathcal{E} . However, this second parameter domain is not connected; its choice seems therefore less natural than the former one.

We recall that a *neighborhood* of a point x in \mathbb{R}^m is any subset V of \mathbb{R}^m for which there exists an open ball $B(x, \varepsilon)$ centered at x , of radius ε (with $\varepsilon > 0$) and such that $B(x, \varepsilon) \subseteq V$.

Definition 4.8 Assume $\mathcal{D} \subseteq \mathbb{R}^m$, for some m in \mathbb{N} . The model $(\mathcal{D}, f, \mathcal{O})$ is *locally identifiable* at a point x in \mathcal{D} if f is injective on some relative neighborhood of x in \mathcal{D} . By a *relative neighborhood* of a point x in \mathcal{D} , we mean the intersection with \mathcal{D} of some neighborhood of x in \mathbb{R}^m . The model is *locally identifiable* if it is locally identifiable at any of its points.

Example 4.9 We take the same probabilistic model as in Example 4.7 (see also Figure 4.1). The model is locally identifiable at any point x in \mathcal{D} except for $x = 0.4$. It is thus not locally identifiable.

Any identifiable model is locally identifiable at every point of its parameter domain (in Definition 4.8, use \mathcal{D} as the relative neighborhood). The converse does not hold, as shown in the next example.

Example 4.10 Assume $m = 1$ and $n = 2$, $\mathcal{D} = [-1.1, 1.1]$ and $\mathcal{O} = [0, 1]^2$. Take the prediction function

$$f : [-1.1, 1.1] \rightarrow [0, 1]^2 : x \mapsto (x^3 - x + 0.5, 0.5x^2 + 0.1). \quad (4.20)$$

Figure 4.2 shows the image of f , that is, the set $f(\mathcal{D})$ of predicted points, as a curve in the outcome space $[0, 1]^2$ (a full square). The resulting model $(\mathcal{D}, f, \mathcal{O})$ is testable, and even quantitatively testable because with μ the measure in the plane we have $\mu(f(\mathcal{D})) = 0$ and $\mu(\mathcal{O}) = 1$. It is identifiable at any parameter point distinct from $x = -1$ and $x = 1$, in other words at any predicted point distinct from $(0.5, 0.6)$. To show this, assume $f(x) = f(x')$ for some distinct x, x' in \mathcal{D} . Then

$$\begin{cases} x^3 - x + 0.5 &= x'^3 - x' + 0.5 \\ 0.5x^2 + 0.1 &= 0.5x'^2 + 0.1 \end{cases} \quad (4.21)$$

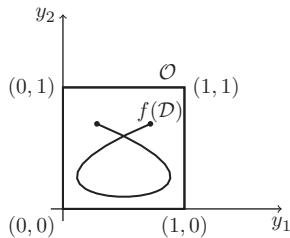


Figure 4.2 A representation of the prediction range $f(\mathcal{D})$ in the outcome space $\mathcal{O} = [0, 1]^2$ as in Example 4.10.

thus

$$\begin{cases} x^3 - x = x'^3 - x' \\ x^2 = x'^2 \end{cases} \quad (4.22)$$

so that $x = \pm x'$. Because we look for $x \neq x'$, we keep only $x = -x' \neq 0$ and then derive

$$\begin{cases} x^3 - x = -x^3 + x \\ x = -x' \end{cases} \quad (4.23)$$

so that $x^3 - x = 0$. This leaves us with $x = -1$ and $x = 1$ (correspondingly, $x' = 1$ and $x' = -1$) as the only parameter points where the model is not identifiable. The model is locally identifiable at any parameter point: this follows from the fact that the mapping $f_2 : [-1.1, 1.1] \rightarrow [0, 1] : x \mapsto x^2 + 0.1$ is locally injective except at 0 (that is, any point of its parameter domain distinct from 0 has a relative neighborhood on which the restriction of f is injective), and at 0 the mapping $f_1 : [-1.1, 1.1] \rightarrow [0, 1] : x \mapsto x^3 - x + 0.5$ is locally injective (that f itself is locally injective follows also from later Proposition 4.21).

We now briefly describe a famous probabilistic model, Luce's choice model, and investigate its testability and identifiability. Luce's choice model is prototypical for models having a measurement-theoretic foundation (Krantz *et al.*, 1971), and constitutes a case of derived measurement (Suppes & Zinnes, 1963; Roberts, 1979). The probability of choosing an element from a set of alternatives is explained on the basis of a numerical characterization of their attractiveness. A representation theorem formulates conditions on the choice probabilities (so-called axioms) that are sufficient for proving the existence of the postulated measurement of attractiveness. The testability of Luce's choice model is related to how restrictive these axioms are. A uniqueness theorem characterizes the scale type of the measurement (Stevens, 1946). Its intimate relation to the model's identifiability is examined below.

Example 4.11 Let \mathcal{A} be a collection of n alternatives. Suppose we observe for each subset X of \mathcal{A} and each alternative i in X the frequency $f_{i,X}$ that individuals choose alternative i as their more preferred one when presented the set X to choose from (exactly one alternative is to be chosen from X). Let \mathbb{R}^ℓ be the space consisting of all vectors y with real components $y_{i,X}$ indexed by pairs i, X with $i \in X \subseteq \mathcal{A}$ and $|X| \geq 2$. The outcome space \mathcal{O} consists of those vectors y in \mathbb{R}^ℓ that satisfy $\sum_{i \in X} y_{i,X} = 1$ for any X subset of \mathcal{A} with $|X| \geq 2$.

A choice model due to Luce (1959) assumes that choices are governed by weights assigned to the alternatives, where for alternative i the strictly⁷ positive weight w_i represents the attractiveness of i . All possible vectors of weights form the parameter domain \mathcal{D} ; in other words, $\mathcal{D} = (\mathbb{R}_{>0})^{\mathcal{A}}$. Here is the essential assumption of the model: for a fixed subset X , the probabilities of choices of the various

⁷ Strict positiveness is assumed here to simplify the exposition, see Luce (1959) for the general case.

alternatives i in X are proportional to the weights w_i . Thus, the prediction function f maps any point w from \mathcal{D} to the outcome point with components $p_{i,X}$ (for $i \in X \subseteq \mathcal{A}$ and $|X| \geq 2$) equal to

$$p_{i,X} = \frac{w_i}{\sum_{j \in X} w_j}. \quad (4.24)$$

Luce (1959) provides an axiomatization (in terms of more general concepts, namely the probability $p_{Y,X}$ that the choice of the best alternative in a given set X falls in the subset Y of X) which formulates conditions sufficient for proving the existence of the weights in (4.24).

As it is presented here, Luce's choice model is not an identifiable model because if we multiply all weights w_i by the same strictly positive real number, the values $p_{i,X}$ remain the same. This means that measurement of attractiveness is on a ratio scale (Stevens, 1946). Having available this characterization of the non-uniqueness of the representation allows for restricting the parameter domain to restore identifiability. Indeed, it is easily checked for two alternatives i, j in the subset X of \mathcal{A} that

$$\frac{w_i}{w_j} = \frac{p_{i,X}}{p_{j,X}}. \quad (4.25)$$

In case $X = \{i, j\}$, we deduce

$$\frac{w_i}{w_j} = \frac{p_{i,\{i,j\}}}{p_{j,\{i,j\}}}. \quad (4.26)$$

Hence, we may fix the weight of one arbitrarily chosen alternative a to 1, so that the weights of any other alternative i must satisfy

$$w_i = \frac{p_{i,\{i,a\}}}{p_{a,\{i,a\}}}. \quad (4.27)$$

Under the assumption $w_a = 1$, the resulting model becomes identifiable.

Another consequence of Equation (4.25) is that Luce's choice model is quantitatively testable as soon as $|\mathcal{A}| \geq 3$. Indeed, if $i, j \in X \subset Y \subseteq \mathcal{A}$, we must have

$$\frac{p_{i,X}}{p_{j,X}} = \frac{p_{i,Y}}{p_{j,Y}}, \quad (4.28)$$

which for $i \neq j$ is the equation of a surface in \mathbb{R}^ℓ containing all the predicted points (the remark holds even if we fix $w_a = 1$ as above).

At the beginning of this section, we stated two fundamental questions about a model $(\mathcal{D}, f, \mathcal{O})$. Assuming now that \mathcal{D} and \mathcal{O} are each embedded in a numerical space, we may add variants of the questions:

1. Is the model testable? Is it quantitatively testable?
2. Is the model identifiable? Is it locally identifiable? At which points is it identifiable?

These questions and their variants will be investigated for two models of student responses in the setting of knowledge structures: in Sections 4.6 and 4.7 we consider the CRM, then in Sections 4.8 to 4.10 we consider the BLIM. We first present in the next section mathematical tools that we will be using, and then in Sections 4.4 and 4.5 the basics of KST. Latent class models will appear in Section 4.11.

4.3 Mathematical Tools

The investigations of the CRM and the BLIM rely on tools from different mathematical disciplines; the main reason is that the prediction function is linear in the CRM, but not in the BLIM. For the first model, we rely on concepts and results from discrete, convex geometry; Subsection 4.3.1 briefly introduces the basics. For the second model, the tools come from mathematical analysis; Subsections 4.3.2 to 4.3.6 are devoted to their description.

4.3.1 Convex Polytopes

Let us briefly recall that a (convex) *polytope* in \mathbb{R}^d is the convex hull of a finite set of points (for background on polytopes, see, for instance, Grünbaum, 2003, or Ziegler, 1998). It is a *0/1-polytope* when all these points can be selected with their coordinates equal to 0 or 1. A polytope in \mathbb{R}^d is *full* if it affinely generates \mathbb{R}^d .

Let P be a polytope in \mathbb{R}^d . An important result states that P always admits a *linear description*, that is, a system of affine⁸ equations and inequalities on \mathbb{R}^d whose set of solutions coincides with P . Conversely, if bounded, the set of solutions of a system of affine equations is a polytope.

A *face* of P is either P itself or the intersection of P with a *supporting hyperplane*, that is, a hyperplane H such that P lies entirely on one closed side of P . The empty set is a face of P . The zero-dimensional faces of P are singleton sets $\{v\}$; by definition, v is then a *vertex* of P . The *edges* of P are the one-dimensional faces, the *facets* of P are the faces of dimension one less than the dimension of P .

Among all subsets of points whose convex hull is a given polytope P , there is a minimum one (w.r.t. inclusion); it is formed by the vertices of P . Assume now that P is full in \mathbb{R}^d ; among all systems of affine equations whose set of solutions equals P , there is one whose equations appear in all such systems. The equations of the minimum system define supporting hyperplanes whose various intersections with P are exactly the facets of P . If P is not full, things are slightly more involved; see Grünbaum (2003) or Ziegler (1998) (anyway, one may replace \mathbb{R}^d with the subspace that is affinely generated by P).

For a given, full polytope, we thus have two minimum descriptions: the first by the set of vertices, the second by the set of facets. In practical applications, the vertices are described by their coordinates in \mathbb{R}^d , and the facets by the coefficients

⁸ An *affine equation* is of the form $a_1 x_1 + a_2 x_2 + \cdots + a_d x_d = b$; it is a *linear equation* when $b = 0$.

in the equations of the hyperplanes defining the facets. There are algorithms to transform one description into the other; of course, their computer implementations run only for limited dimensions and numbers of vertices/facets. One reason is that for many families of polytopes with unbounded dimensions, the number of facets is exponential in the number of vertices (and for other families the other way round).

4.3.2 Two Theorems (and More Propositions) from Mathematical Analysis

A central concept in approaching quantitative testability and local identifiability of a model at a parameter point x is that of the differential at x of the prediction function f or, equivalently, the *Jacobian matrix* at x of f . The rank of the differential, which is the same as the rank of the Jacobian matrix, plays a crucial role, as will be shown in this section.

In the present subsection, we introduce the concepts from a mathematical point of view and state a few easy facts on the rank of the Jacobian matrix. We also mention two fundamental results: the rank theorem, and Sard's theorem. Some general references for (most of) the mathematics we review here are Boothby (1975), Lang (1997), Rudin (1976) and, at a higher level, Dieudonné (1960a,b), and Lang (1993).

After some comments on probabilistic models, in the next subsection we derive consequences for quantitative testability and local identifiability. The section ends with a summary of the tools that we described for the investigation of probabilistic models.

All along this subsection, we let \mathcal{E} be an open subset of \mathbb{R}^m and $f : \mathcal{E} \rightarrow \mathbb{R}^n$ be a function. The Jacobian matrix of the function f evaluated at the point x° in \mathcal{E} is the $n \times m$ matrix of partial derivatives (if they exist)

$$\mathbf{J}_f(x^\circ) = \left(\frac{\partial f_i}{\partial x_j}(x^\circ) \right) \quad i = 1, \dots, n, j = 1, \dots, m. \quad (4.29)$$

The columns of the Jacobian matrix are indexed as the components x_j , $j = 1, 2, \dots, m$, of the vector x in \mathbb{R}^m ; the rows as the components of the image $f(x)$ in \mathbb{R}^n . The importance of the Jacobian matrix comes from its strong relationship with the differential. By definition the *differential* $df(x^\circ)$ of f at x° , if it exists, is a linear mapping from \mathbb{R}^m to \mathbb{R}^n such that the function f is well-approximated near x° by the affine function

$$\alpha : \mathbb{R}^m \rightarrow \mathbb{R}^n : x \mapsto \alpha(x) = f(x^\circ) + df(x^\circ)(x - x^\circ). \quad (4.30)$$

More precisely, there must be some neighborhood V of x° , contained in \mathcal{E} , and some function $\varepsilon : V \rightarrow \mathbb{R}$ such that, first, $\varepsilon(x)$ tends to 0 when x tends to x° , and second, for any point x in V

$$|f(x) - \alpha(x)| \leq |x - x^\circ| \cdot \varepsilon(x). \quad (4.31)$$

When the differential $df(x^\circ)$ exists, we say that f is *differentiable* at x° . In this case, the linear mapping $df(x^\circ)$ is represented in the canonical bases of \mathbb{R}^n and \mathbb{R}^m

by a matrix, which happens to be the Jacobian matrix. In other words, if we represent points like x , x° , $f(x)$, etc., by column vectors containing their coordinates, we get

$$df(x^\circ)(x - x^\circ) = J_f(x^\circ)(x - x^\circ). \quad (4.32)$$

Thus, when f is differentiable at x° , all the partial derivatives $\frac{\partial f_i}{\partial x_j}(x^\circ)$ exist. However, the converse implication does not hold (see, however, Proposition 4.12 below).

The range of the affine approximation α is the affine subspace $f(x^\circ) + df(x^\circ)(\mathbb{R}^m)$. It is obtained by taking the linear subspace spanned by the column vectors of the Jacobian matrix, and by translating this subspace by the vector $f(x^\circ)$. The dimension of the range equals by definition the *rank* of the differential $df(x^\circ)(x)$. It is also the *rank* of the Jacobian matrix $J_f(x^\circ)$, denoted by $\text{rk}(J_f(x^\circ))$. The latter equals the maximum number of linearly independent columns (or, equivalently, rows) of $J_f(x^\circ)$; it is also the largest number of rows (or, equivalently, columns) of a square submatrix of $J_f(x^\circ)$ with non-zero determinant. The Jacobian matrix $J_f(x^\circ)$ being a $n \times m$ matrix, its rank cannot be greater than n or m . If $\text{rk}(J_f(x^\circ)) = \min\{n, m\}$ then the Jacobian matrix J_f is said to have *full rank* at x° , otherwise it is *rank-deficient*.

Under some conditions, knowing the rank of the Jacobian matrix on an open neighborhood of x° allows for inferring whether the function is *locally injective* at x° (meaning injective on some neighborhood of x° : this is just the mathematical synonym for locally identifiable). For these kinds of inferences the prediction function f has to be sufficiently “smooth.” In particular, the function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is of *class C^s* or in short C^s at a point x° (of \mathcal{E}) if all partial derivatives of f up to order s exist in some open neighborhood of x° , and they all are continuous on the neighborhood. If f is C^s at all points of \mathcal{E} , it is of *class C^s* , or C^s . The function is (of *class*) C^∞ if it is C^s for all $s \geq 1$. The function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is *analytic* or C^ω if for each point x° in \mathcal{E} there is an open neighborhood V of x° included in \mathcal{E} such that for each x in V , the Taylor series expansion of f around x° , when evaluated at x , converges to $f(x)$ (for details, see for instance Dieudonné, 1960a). Any analytic function is C^∞ . For applications of the mathematical tools to various models (for instance, the BLIM), it is crucial that polynomial functions are analytic.

Proposition 4.12 *If the function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is C^1 at the point x° , it is differentiable at x° and even at all points of some open neighborhood of x° included in \mathcal{E} .*

So when the function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is C^1 at x° , we may speak of the rank of the Jacobian matrix at all points in some neighborhood of x° . Moreover, the following then holds.

Proposition 4.13 *Assume the function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is C^1 at the point x° . For all points x in some open neighborhood of x° included in \mathcal{E} , we have $\text{rk}(J_f(x^\circ)) \leq \text{rk}(J_f(x))$.*

Indeed, if the Jacobian matrix $J_f(x^\circ)$ has rank k , some of its $k \times k$ submatrices have non-zero determinants. By continuity of the submatrix elements and thus of the determinant, the same submatrix has non-zero determinants at all the points x in some neighborhood of x° , thus $\text{rk}(J_f(x^\circ)) \leq \text{rk}(J_f(x))$. One cannot ensure that $J_f(x)$ be of rank exactly k , as the function $\mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^2$ shows at $x^\circ = 0$ (with $k = 0$).

Here are two easy consequences of Proposition 4.13. Assume again $f : \mathcal{E} \rightarrow \mathbb{R}^n$, where \mathcal{E} is an open subset of \mathbb{R}^m and $x^\circ \in \mathcal{E}$. If f is C^1 and $J_f(x^\circ)$ has rank m (resp. n), then $J_f(x)$ has rank m (resp. n) for all points x in some open neighborhood of x° .

A point x° in \mathcal{E} is *regular* for the function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ if on some open neighborhood of x° , included in \mathcal{E} , the function f is C^1 and moreover the rank of the Jacobian matrix is constant on the neighborhood. For an analytic function, almost all points are regular: the following statement follows, for example, from Fisher (1966, theorem 5.A.1).

Proposition 4.14 *Let $f : \mathcal{E} \rightarrow \mathbb{R}^n$ be an analytic function defined on the open subset \mathcal{E} of \mathbb{R}^m . Denoting by G the set of regular points of f , we have $\lambda(\mathcal{E} \setminus G) = 0$ where λ is the (Lebesgue) measure in \mathbb{R}^m .*

By the definition of a regular point, the set of all regular points is necessarily open. The following, related result appears as proposition B.4 in Bamber and van Santen (1985).

Proposition 4.15 *Let $f : \mathcal{E} \rightarrow \mathbb{R}^n$ be an analytic function defined on the connected, open subset \mathcal{E} of \mathbb{R}^m . Denote by M the subset of \mathcal{E} consisting of the points where $\text{rk}(J_f)$ attains its maximum. Then M is open, and moreover $\lambda(\mathcal{E} \setminus M) = 0$.*

Bamber and van Santen (1985) give also a probabilistic extension (in their Proposition B.5).

Proposition 4.16 *As in the previous proposition, let $f : \mathcal{E} \rightarrow \mathbb{R}^n$ be an analytic function defined on the connected, open subset \mathcal{E} of \mathbb{R}^m . Moreover, suppose \mathcal{D} is a subset of \mathcal{E} with $\lambda(\mathcal{D}) > 0$ and X is a random point in \mathcal{D} with a probability density function. Then*

$$\mathbb{P}\left(\text{rk}(J_f(X)) = \max_{x \in \mathcal{D}} \text{rk}(J_f(x))\right) = 1. \quad (4.33)$$

Here is now our first crucial theorem from classical, mathematical analysis; it is often called the (constant) rank theorem, as in Boothby (1975, theorem 7.1) or Rudin (1976, theorem 9.32). In rough terms it says that, close to a regular point, any function looks like a linear projection. For a natural number q with $q \geq 1$, a C^q diffeomorphism g defined on an open subset V of \mathbb{R}^m is an invertible C^q mapping from V to some open subset W of \mathbb{R}^m such that its inverse is also a C^q mapping (from W to V). The theorem is illustrated in Figure 4.3.

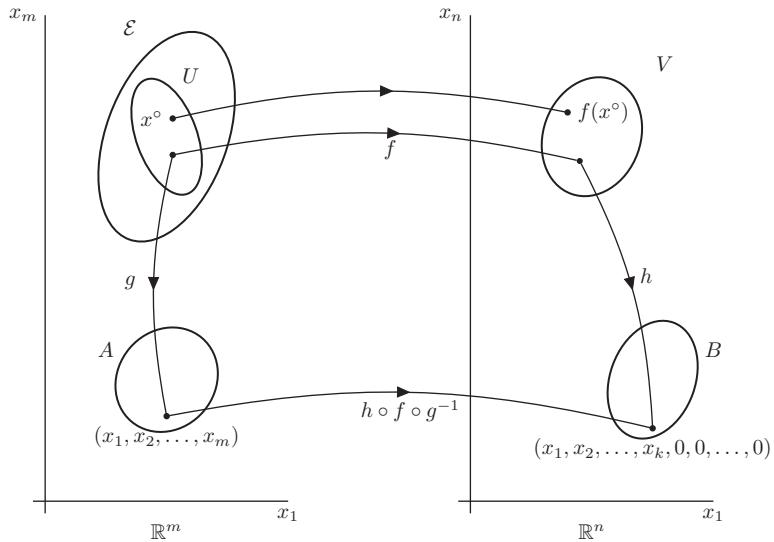


Figure 4.3 An illustration of Theorem 4.17.

Theorem 4.17 (Rank theorem) Let \mathcal{E} be an open subset of \mathbb{R}^m , $f : \mathcal{E} \rightarrow \mathbb{R}^n$ be a function and x° be a point of \mathcal{E} . Assume that f is C^q at x° for some q in $(\mathbb{N} \setminus 0) \cup \{\infty, \omega\}$ and moreover that x° is a regular point with $\text{rk}(J_f(x^\circ)) = k$. Then there exists some open neighborhood U of x° included in \mathcal{E} , some open neighborhood V of $f(x^\circ)$ and C^q diffeomorphisms $g : U \rightarrow A$ and $h : V \rightarrow B$ (where A and B are open subsets of, respectively, \mathbb{R}^m and \mathbb{R}^n) such that for any point (x_1, x_2, \dots, x_m) in A there holds

$$(h \circ f \circ g^{-1})(x_1, x_2, \dots, x_m) = (x_1, x_2, \dots, x_k, 0, 0, \dots, 0). \quad (4.34)$$

In the last statement, it must be understood that the number of zeros shown in $(x_1, x_2, \dots, x_k, 0, 0, \dots, 0)$ equals $n - k$.

There is more to infer from Theorem 4.17 (see Figure 4.4 for $k = m - 1$). To this aim, assume that A is an open ball centered at $g(x^\circ)$ (this is not a restriction). The points z in A that have the same image by $h \circ f \circ g^{-1}$ as the point $g(x^\circ)$ are exactly those points in A whose first k coordinates coincide with those of $g(x^\circ)$. These points constitute the intersection I of the ball A with an affine subspace of dimension $m - k$; thus, I is an $(m - k)$ -dimensional open ball in this subspace. By pulling the $(m - k)$ -dimensional ball I back by the diffeomorphism g^{-1} , we get a subset $g^{-1}(I)$ of U (and of \mathcal{E}) containing x° on which the function f is constant. By construction, $g^{-1}(I)$ is a (piece of) a surface in \mathbb{R}^m (it is even a connected “ C^q -manifold without boundary” of dimension $m - k$).

The second crucial result we quote from analysis is a more advanced theorem due to Sard (1942, 1958). Assume that the function $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is differentiable. A point x in \mathcal{E} is *critical of rank k* if $\text{rk}(J_f(x)) = k < n$, and *critical* if it is critical of rank k for some k (with thus $k < n$). A simplified version of Sard’s theorem

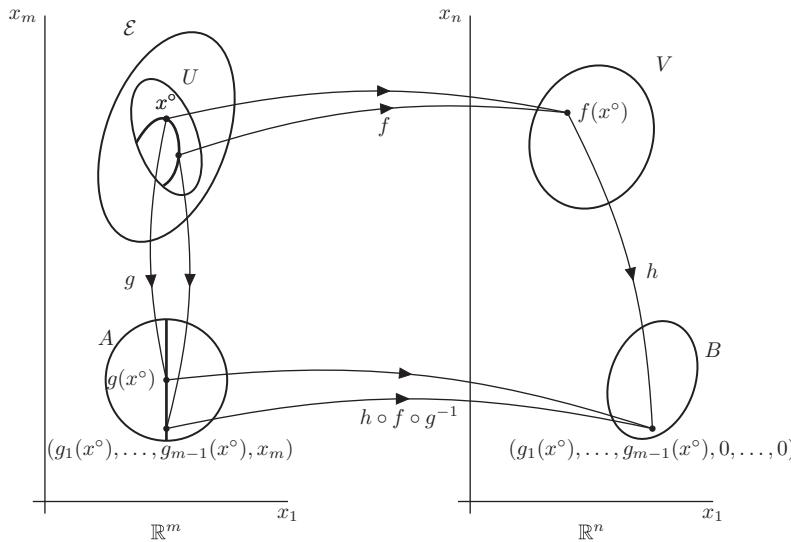


Figure 4.4 An illustration of the construction of the surface M (for $k = 1$, a curve which is the pullback of an open, vertical segment).

states that when f is C^∞ , the images by f of all singular points form a subset of null measure in \mathbb{R}^n (see, for instance, Dieudonné, 1960b, section 16.23.1). In case $m < n$, the conclusion is that $f(\mathcal{E})$ has null measure (because then all points x in \mathcal{E} are critical). In case $m \geq n$, we may have $\mu(f(\mathcal{E})) > 0$ but Sard's theorem says that the images of critical points (sometimes called “critical values”) form a null-measure set. The version of Sard's theorem that we now state is more elaborate (the smoothness requirements being weaker).

Theorem 4.18 (Sard's theorem) *Assume $f : \mathcal{E} \rightarrow \mathbb{R}^n$ is differentiable on the open subset \mathcal{E} and let k be any value in $\{0, 1, \dots, n - 1\}$. If f is C^q for some $q \geq \max\{1, (m - k)/(n - k)\}$, then the images of all critical points of rank at most k form a subset of \mathbb{R}^n having null measure.*

The next two subsections derive some consequences for probabilistic models from the mathematical results we just presented.

4.3.3 Investigating Probabilistic Models

From now on in this section, we consider a model $(\mathcal{D}, f, \mathcal{O})$, with $\mathcal{D} \subseteq \mathbb{R}^m$ and $\mathcal{O} \subseteq \mathbb{R}^n$. To use results from mathematical analysis, we often assume that \mathcal{D} has a non-empty interior.

As illustrated in Examples 4.4 and 4.5, the CRM and the BLIM have a common parameter domain \mathcal{D} with empty interior. However, in the topology of the affine subspace generated by \mathcal{D} , most points of \mathcal{D} are interior points. In practice, it suffices to eliminate one of the parameters π_K by taking into account $\sum_{K \in \mathcal{K}} \pi_K = 1$

to get a parameter domain with non-empty interior. The general construction is detailed for the BLIM in Section 4.8.1.

At any interior point x of the parameter domain \mathcal{D} where the prediction function is differentiable, we have the Jacobian matrix $J_f(x)$. In the next two sections, we show how to exploit the value of the rank of this matrix to derive information on the model.

4.3.4 Results about Model Testability

For quantitative testability (in the sense of Definition 4.6), we start with a negative conclusion for our model $(\mathcal{D}, f, \mathcal{O})$ (recall $\mathcal{D} \subseteq \mathbb{R}^m$ and $\mathcal{O} \subseteq \mathbb{R}^n$).

Proposition 4.19 *Assume that f is C^1 at some point x° in the interior of \mathcal{D} . If the Jacobian matrix $J_f(x^\circ)$ has rank n , the model cannot be quantitatively testable.*

By Proposition 4.13, the point x° must be regular with rank n ; there is some open neighborhood \mathcal{E} of x° contained in U on which f is C^1 and the rank is constant. Using the notation of the rank theorem (Theorem 4.17), we infer for $k = n$ that the mapping $h \circ f \circ g^{-1}$ is surjective from A to B . Hence $f(U)$ includes the open subset V of \mathbb{R}^n , so that $f(\mathcal{D})$ cannot have zero measure.

To reach a positive conclusion on quantitative testability, we are tempted to assume that the rank of the Jacobian matrix is everywhere less than n . However, we need to be careful about the (possible) points of the parameter domain \mathcal{D} which lie on the boundary. For this reason, \mathcal{D} is enlarged into an open subset \mathcal{E} of \mathbb{R}^m , and the prediction function is extended to \mathcal{E} . By relying on Sard's theorem (here Theorem 4.18), Bamber and van Santen (1985) thus obtained the following result in their theorem 9.

Proposition 4.20 *Suppose \mathcal{E} is an open subset of \mathbb{R}^m and the differentiable function f maps \mathcal{E} into \mathbb{R}^n . Suppose $\mathcal{D} \subseteq \mathcal{E}$ and that $(\mathcal{D}, f, \mathcal{O})$ is our model, with $\mu(\mathcal{O}) > 0$ (where we have restricted the function f to \mathcal{D}). Let $r = \max_{x \in \mathcal{D}} \text{rk}(J_f(x))$. If $r < n$ and iff f is C^q on \mathcal{E} for some $q \geq \max\{1, (m - r)/(n - r)\}$, then the model $(\mathcal{D}, f, \mathcal{O})$ is quantitatively testable.*

By the assumptions in Proposition 4.20, any point in \mathcal{D} is critical for the function f defined on \mathcal{E} . Hence, $f(\mathcal{D})$ is contained in the set of images of critical points. Theorem 4.18 asserts that this set has null measure, and so does also the predictive range $f(\mathcal{D})$.

4.3.5 Results about Model Identifiability

We first reach a positive conclusion for the identifiability of our model $(\mathcal{D}, f, \mathcal{O})$ (noted for instance by Smith and Minda, 1998, appendix A).

Proposition 4.21 *Assume that the prediction function f is C^1 at some interior point x° of the parameter domain \mathcal{D} , and moreover that the Jacobian matrix $J_f(x^\circ)$ has rank m . Then the model is locally identifiable at x° .*

Here is an argument: by Proposition 4.13, there exists some neighborhood V of x° contained in the interior of \mathcal{D} such that $J_f(x)$ has rank m at each point of V . Hence Proposition 4.21 follows at once from Theorem 4.17. By providing a counter-example in the context of “factor models,” Shapiro and Browne (1983) demonstrate that the converse implication need not be true in general. A very simple counter-example comes from the model $(\mathbb{R}, f, \mathbb{R})$ with $f(x) = x^3$. Local identifiability holds at $x^\circ = 0$, although the rank of the Jacobian at 0 is 0, strictly less than $n = 1$.

As we now indicate, the implication converse to the one in Proposition 4.21 holds for regular points (Wald, 1950).

Proposition 4.22 *Assume that the prediction function f is C^1 and regular at some interior point x° of the parameter domain \mathcal{D} . Then the model is locally identifiable at x° if and only if the Jacobian matrix $J_f(x^\circ)$ has rank m .*

Let $k = \text{rk}(J_f(x^\circ))$. If $k = m$, Proposition 4.21 applies. If $k < m$, Theorem 4.17 entails that we cannot get local identifiability at x° : in its notation, $h \circ f \circ g^{-1}$ is constant on the set Z of points $\{(0, 0, \dots, 0, x_m)\}$ lying in A . Hence f is constant on $g^{-1}(Z)$.

Thus, $k < m$ entails that the model is not identifiable. As explained after Definition 4.3, it is always possible to restrict the parameter domain in order to get an identifiable model without modifying the outcome space. In the present setting, the construction would replace \mathcal{D} with a subset \mathcal{C} satisfying at least the following condition. For each point x° in \mathcal{C} , consider in the notation of page 144 the piece of surface $g^{-1}(I)$ (which contains x°). Then \mathcal{C} must in particular meet each of these pieces in at most one point.

We conclude this subsection with another result of Bamber and van Santen (1985, proposition 20).

Proposition 4.23 *Assume that the prediction function f is C^1 on some open subset U of \mathcal{D} , and that the Jacobian matrix $J_f(x)$ has rank strictly less than m at every point x of U . Then the model $(\mathcal{D}, f, \mathcal{O})$ is not locally identifiable.*

If $J_f(x)$ attains its maximum at a point x° of U , the maximum is also attained at every point in some neighborhood of x° contained in U ; in other words, x° is regular. We conclude the argument by referring to Proposition 4.22.

Corollary 4.24 *Suppose the prediction function f is C^1 on some open subset U of \mathcal{D} . If $m > n$ the model is not locally identifiable.*

4.3.6 Summary

To investigate local identifiability of a probabilistic model $(\mathcal{D}, f, \mathcal{O})$ by applying the mathematical tools we just reviewed, we need to know the rank of the Jacobian matrix at points of the interior of the parameter domain \mathcal{D} . If we can determine the rank in an algebraic way, we then look for the maximum rank that is attained at some point(s) in the interior. Then we can immediately draw upon the above results.

Table 4.1 *Summary of implications of the model $(\mathcal{D}, f, \mathcal{O})$, where $\mathcal{D} \subseteq \mathbb{R}^m$ and $\mathcal{O} \subseteq \mathbb{R}^n$; moreover, r denotes the maximum value of $\text{rk}(J_f)$ on the interior of \mathcal{D} (or on the fourth line, on \mathcal{D} itself).*

Thm.	Assumption on f	Rank assumption	Conclusion
4.19	C^1 at x°	$\text{rk}(J_f(x^\circ)) = n$	not quantit. testable
4.19	C^1 on the interior	$r = n$	not quantit. testable
4.20	see theorem	$r < n$	quantit. testable
4.20	see theorem	$m < n$	quantit. testable
4.21	C^1 at x°	$\text{rk}(J_f(x^\circ)) = m$	loc. identifiable at x°
4.22	regular at x°	$\text{rk}(J_f(x^\circ)) < m$	not loc. ident. at x°
4.23	C^1 on the interior	$r < m$	not loc. ident.
4.24	C^1 on some open subset	$m > n$	not loc. ident.

If the rank of the Jacobian matrix can only be determined numerically at particular points in the parameter space, then *a priori* we cannot be sure to have identified its maximum rank. However, Proposition 4.16 tells us that, loosely speaking, if the prediction function is analytic then the Jacobian matrix has maximum rank almost everywhere in \mathcal{D} . This opens up an avenue to numerically computing the maximum rank of the Jacobian matrix. Randomly selecting several points from the parameter space (e.g., with respect to a uniform distribution on \mathcal{D}), and evaluating the Jacobian matrix at these points, one can be almost sure to detect its maximum rank. Then the tools of this section are applicable.

We summarize in Table 4.1 some conclusions about the model $(\mathcal{D}, f, \mathcal{O})$. Here, “interior” always designates the interior of \mathcal{D} , “open subset” means “open subset of \mathcal{D} ” and x° designates some point in the interior. Moreover, under the assumption that the prediction function f is differentiable on the interior, we denote by r the maximum value taken by the rank of the Jacobian of f at the interior points. Remember that a model which is not testable is also not quantitatively testable; if it is not locally identifiable, it is also not identifiable.

4.4 Knowledge Structures: Combinatorial Aspects

Both the CRM and the BLIM are probabilistic models based on a “knowledge structure.” We recall in this section the basic combinatorial ingredients of knowledge structure theory (KST), and in the next section the probabilistic extension of the theory. The reader looking for more motivations, details and results is referred to the chapter by Doignon and Falmagne (2016) in the first volume of this Handbook.

Definition 4.25 A *knowledge structure* is a pair (Q, \mathcal{K}) consisting of a non-empty, finite set Q and a collection \mathcal{K} of subsets of Q such that $\emptyset \in \mathcal{K}$ and $Q \in \mathcal{K}$. Its

domain is the set Q , its *items* are the elements of Q , and its *knowledge states* are the elements of \mathcal{K} .

For any item q in Q , we denote with \mathcal{K}_q the subcollection of \mathcal{K} consisting of all the states containing q , that is $\mathcal{K}_q = \{K \in \mathcal{K} \mid q \in K\}$. A knowledge structure (Q, \mathcal{K}) is *discriminative* when for any two items q and r in the domain, the equality $\mathcal{K}_q = \mathcal{K}_r$ holds only if $q = r$.

We often abbreviate (Q, \mathcal{K}) into \mathcal{K} (notice $Q = \cup \mathcal{K}$). The following notation is also useful: $\mathcal{K}_{\bar{q}} = \{K \in \mathcal{K} \mid q \notin K\}$ (thus \mathcal{K}_q and $\mathcal{K}_{\bar{q}}$ partition \mathcal{K}).

Example 4.26 Here are two examples of knowledge structures on the same domain $Q = \{a, b, c, d\}$:

$$\begin{aligned}\mathcal{K}^{(1)} &= \{\emptyset, \{a\}, \{b\}, \{a, c\}, \{b, d\}, Q\}, \\ \mathcal{K}^{(2)} &= \{\emptyset, \{a\}, \{c\}, \{a, b\}, \{a, c\}, \{c, d\}, \{a, b, c\}, \{a, c, d\}, Q\}.\end{aligned}$$

Both knowledge structures are discriminative. They are sketched in Figure 4.5.

The knowledge structure $\mathcal{K}^{(1)}$ in Example 4.26 has two possibly unpleasant features. First, assume a student is in a knowledge state distinct from the whole domain Q . To learn all the remaining items in Q , she will need to acquire two distinct items at the same time: indeed, there is no knowledge state containing all of the items but one. Second, say a student masters only item a . Then he cannot learn next the single additional item b because the subset $\{a, b\}$ is not a state. This is strange, because he could have learned item b when they were in the empty state of knowledge. The next definition explicitly rules out these two features.

Definition 4.27 A *learning space* (Q, \mathcal{K}) is a knowledge structure which satisfies the following two conditions:

1. For any non-empty state K , there exists some item q in K such that $K \setminus \{q\}$ is also a state.
2. If K is a state and q, r are two items such that $K \cup \{q\}$ and $K \cup \{r\}$ are states, then $K \cup \{q, r\}$ is also a state.

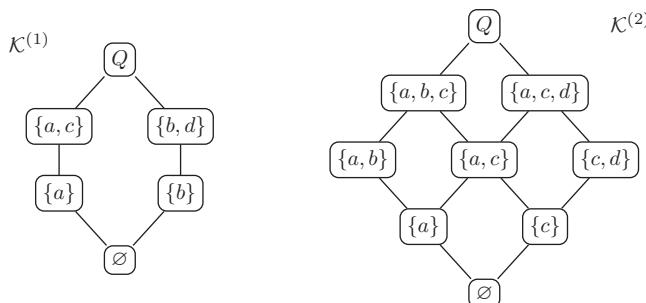


Figure 4.5 The two examples of knowledge structures in Example 4.26.

In Example 4.26 (see also Figure 4.5), the knowledge structure $\mathcal{K}^{(2)}$ is a learning space, but $\mathcal{K}^{(1)}$ is not. There are many equivalent characterizations of learning spaces. One of them (in Proposition 4.30 below) is based on concepts introduced in the following two definitions.

Definition 4.28 A *knowledge space* \mathcal{K} is a knowledge structure which is closed under union, or *union closed* (\cup -closed), that is, $\cup\mathcal{C} \in \mathcal{K}$ for any subcollection \mathcal{C} of \mathcal{K} .

For two sets K and L , their *symmetric difference* is $K \Delta L = (K \setminus L) \cup (L \setminus K)$. For a finite set K , its *cardinality* or number of elements is denoted by $|K|$.

Definition 4.29 A knowledge structure \mathcal{K} is *well-graded* if for any two states K and L in \mathcal{K} with $|K \Delta L| = \ell$, there exists a finite sequence of states $K = K_0, K_1, \dots, K_\ell = L$ such that $|K_{i-1} \Delta K_i| = 1$ for $1 \leq i \leq \ell$.

Proposition 4.30 A knowledge structure is a learning space if and only if it is a well-graded knowledge space.

An important case of a knowledge space arises when the collection of states is moreover closed under intersection.

Definition 4.31 A *quasi-ordinal space* is a knowledge space closed under intersection. A *(partially) ordinal space* is a quasi-ordinal space which is discriminative.

The names for the structures introduced in the last definition come from the relationship with quasi-ordered or partially ordered sets, as established by the next theorem due to Birkhoff (1937).

Definition 4.32 Given a *quasi-order* R on the finite set Q (that is, a reflexive and transitive relation on Q), a *beginning set* of R is any subset B of Q such that, for all p, q in Q

$$\text{if } p R q \text{ and } q \in B, \text{ then } p \in B. \quad (4.35)$$

Because a *partial order* is by definition an antisymmetric quasi-order, we may speak also of beginning sets of partial orders. A *weak order* is a quasi-order which is complete.

Theorem 4.33 (Birkhoff, 1937) Let Q be a finite set. The beginning sets of any quasi-order R on Q form a quasi-ordinal knowledge space on Q . The converse holds, in the sense that any quasi-ordinal knowledge space on Q is obtained in this way from some quasi-order on Q . There results a one-to-one correspondence between the collection of all quasi-orders on Q and the collection of all quasi-ordinal knowledge spaces on Q . A quasi-ordinal knowledge space is a learning space (resp. a chain) exactly if its corresponding quasi-order is a partial order (resp. a weak order).

To make explicit the one-to-one correspondence appearing in Theorem 4.33, let \preccurlyeq be a quasi-order on Q , and \mathcal{K} be the corresponding quasi-ordinal structure on Q . Then \mathcal{K} consists of the beginning sets of \preccurlyeq ; conversely, for any two items p and q in Q , we have $p \preccurlyeq q$ if and only if $\mathcal{K}_p \supseteq \mathcal{K}_q$. Here are for further use a few more notions and elementary properties related to quasi-orders.

Definition 4.34 Let R be a quasi-order on the finite set Q . The relation $R \cap R^{-1}$ is the *equivalence relation* of R . Here, for any p, q in Q , we have $p R \cap R^{-1} q$ if and only if both $p R q$ and $q R p$.

Proposition 4.35 *The equivalence relation of a quasi-order on the finite set Q is indeed an equivalence relation on Q .*

Definition 4.36 Let R and S be two quasi-orders on the finite set Q . Then S is an *extension* of R when $R \subseteq S$, that is, $p R q$ implies $p S q$, for all p, q in Q . It is a *tight extension* if, moreover, R and S have the same equivalence relation; S is a *firm extension* of R if it is a tight extension of R and at the same time a weak order on Q .

Proposition 4.37 *Any quasi-order on a finite set Q has at least one firm extension.*

There are two special and closely related properties that are transversal to a number of different classes of knowledge structures such as, to give some examples, knowledge spaces, learning spaces, and “closure spaces.” They are known as “backward-gradedness” and “forward-gradedness” (Spoto, Stefanutti, & Vidotto, 2012). As will be shown in Section 4.10, these two properties have strong connections with many of the identifiability issues of probabilistic knowledge structures.

Definition 4.38 A knowledge structure (Q, \mathcal{K}) is *backward-graded in item $q \in Q$* if for every state K , the subset $K \setminus \{q\}$ is again a state. It is *backward-graded* if it is backward-graded in some of its items.

A knowledge structure (Q, \mathcal{K}) is *forward-graded in item q* if for every state K , the subset $K \cup \{q\}$ is again a state. It is *forward-graded* if it is forward-graded in some of its items.

Example 4.39 Consider the knowledge structures

$$\begin{aligned}\mathcal{K}^{(3)} &= \{\emptyset, \{a, b\}, Q\}, \\ \mathcal{K}^{(4)} &= \{\emptyset, \{a, b\}, \{a, c\}, Q\}\end{aligned}$$

on domain $Q = \{a, b, c\}$, which are illustrated in Figure 4.6. Then $\mathcal{K}^{(3)}$ is backward-graded in item c , because $\mathcal{K}_c = \{Q\}$, and $Q \setminus \{c\} = \{a, b\} \in \mathcal{K}^{(3)}$. The knowledge structure $\mathcal{K}^{(4)}$ is neither backward-graded nor forward-graded.

Definition 4.40 The *dual* of a knowledge structure (Q, \mathcal{K}) is the structure (Q, \mathcal{K}^Δ) with $\mathcal{K}^\Delta = \{Q \setminus K \mid K \in \mathcal{K}\}$. *Closure spaces* (Q, \mathcal{K}) are the structures whose collection \mathcal{K} of states is closed under intersection.

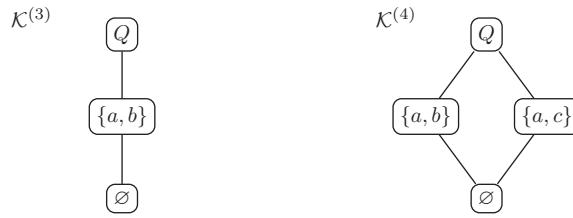


Figure 4.6 The knowledge structures $\mathcal{K}^{(3)}$ and $\mathcal{K}^{(4)}$ of Example 4.39 on the domain $Q = \{a, b, c\}$.

Closure spaces often arise in the area of the so-called competence-based knowledge space theory (Heller, Ünlü, & Albert, 2013) and in that of the cognitive diagnostic models (Bolt, 2007; DeCarlo, 2011; DiBello & Stout, 2007; Haertel, 1984, 1989; Junker & Sijtsma, 2001; Tatsuoka, 2002; Templin & Henson, 2006) as observed by Heller *et al.* (2015).

Closure spaces are thus dual to knowledge spaces. Also, forward-graded knowledge structures are dual to backward-graded structures; more precisely, a structure \mathcal{K} is backward-graded in an item q if and only if \mathcal{K}^Δ is forward-graded in q . Such an interesting property allows easy extension of theoretical results from one of the two families to the dual one, by making the appropriate substitutions.

Many of the most commonly applied knowledge structures happen to be either forward- or backward-graded in some item (some are even both). To begin with, any knowledge space \mathcal{K} (Definition 4.28) which contains at least one singleton $\{q\}$ is forward-graded in q . This happens because \mathcal{K} is closed under union, thus $\{q\} \in \mathcal{K}$ entails $K \cup \{q\} \in \mathcal{K}$ for all $K \in \mathcal{K}$. Moreover, since the dual of a knowledge space is a closure space, it follows at once that every closure space containing $Q \setminus \{q\}$ for some $q \in Q$ is backward-graded in q .

Because of their pedagogically appealing properties, learning spaces (Definition 4.27) represent a rather important class of knowledge structures. They have been successfully applied in existing computerized knowledge assessment systems (such as ALEKS, see Doignon & Falmagne, 2016, and its references). Since learning spaces are well-graded knowledge spaces (Proposition 4.30), they are always forward-graded in at least one item (because of well-gradedness, they always contain at least one singleton $\{q\}$).

Ordinal knowledge spaces (Definition 4.31) are closed under both union and intersection. They are learning spaces, so they are forward-graded. As they form a family closed under duality, they are also backward-graded.

4.5 Probabilistic Knowledge Structures

As mentioned in Section 4.1, we let the knowledge state of a student vary according to some probability distribution, and also take into account careless errors and lucky guesses.

Definition 4.41 A *probabilistic knowledge structure* (Q, \mathcal{K}, π) consists, first, of a knowledge structure (Q, \mathcal{K}) and, second, a probability distribution π on the collection \mathcal{K} of states.

In Definition 4.41, π_K equals the probability that a student's knowledge state be K . Thus the numbers π_K , for $K \in \mathcal{K}$, satisfy $0 \leq \pi_K \leq 1$, and moreover $\sum_{K \in \mathcal{K}} \pi_K = 1$.

Definition 4.42 A *parameterized probabilistic knowledge structure* $(Q, \mathcal{K}, \pi, \beta, \eta)$ is a probabilistic knowledge structure (Q, \mathcal{K}, π) with moreover two numbers β_q and η_q attached to any item q ; these numbers represent, respectively, the *careless error probability* for item q , and the *lucky guess probability* for item q . The notation β and η designate the vectors formed by those numbers, with components indexed by q from Q .

We next give names to parameterized probabilistic knowledge structures in which lucky guesses are assumed to be stochastically impossible, and also to those where careless errors are stochastically impossible.

Definition 4.43 A parameterized probabilistic knowledge structure $(Q, \mathcal{K}, \pi, \beta, \eta)$ is *fair* if $\eta_q = 0$ for all items q . It is *straight* if $\beta_q = 0$ for all items q .

We remark that in Definitions 4.41 and 4.42, the parameters π_K , β_q and η_q have assigned values. In the next section, when we define the CRM, their values vary in a specific range.

4.6 The Correct Response Model: Testability

The CRM is a model providing a possible explanation for the correctness of answers provided by students to each item of a test. Each instance of such a model is based on a particular knowledge structure, whose domain is the set of test items. The next definition generalizes the particular instance of the CRM we met in Section 4.1.

Definition 4.44 Let (Q, \mathcal{K}) be a knowledge structure. The *correct response model* CRM(\mathcal{K}) based on this structure is a probabilistic model whose constituents are as follows. A parameter point has as coordinates the following numbers: for some probability distribution π on \mathcal{K} , all the numbers π_K ; for any item q in Q , the probabilities β_q of a careless error and η_q of a lucky guess. The latter parameter point is conveniently designated as (π, β, η) . Any outcome point has one coordinate $\tau(q)$ for each item q , with $0 \leq \tau(q) \leq 1$, where $\tau(q)$ stands for the probability of a correct answer to item q ; it is denoted shortly as τ . The prediction function maps the parameter point (π, β, η) to the predicted point τ whose coordinate $\tau(q)$, for any item q , is given by

$$\tau(q) = \sum_{K \in \mathcal{K}_q} (1 - \beta_q) \cdot \pi_K + \sum_{K \in \mathcal{K}_{\bar{q}}} \eta_q \cdot \pi_K. \quad (4.36)$$

In Definition 4.44, a parameter point (π, β, η) always satisfies $\pi_K \geq 0$ with $\sum_{K \in \mathcal{K}} \pi_K = 1$, also $0 \leq \beta_q \leq 1$ and $0 \leq \eta_q \leq 1$. Points $(\pi_K)_{K \in \mathcal{K}}$ belong to the real linear space⁹ $\mathbb{R}^{\mathcal{K}}$, which has a basis whose vectors e_K are in bijection with the elements of \mathcal{K} . The real values π_K are the components with respect to that basis of some points, which form the region specified by

$$x_K \geq 0, \quad \sum_{K \in \mathcal{K}} x_K = 1. \quad (4.37)$$

This region is the *simplex* $\Lambda_{\mathcal{K}}$ having as vertices the basis vectors e_K of $\mathbb{R}^{\mathcal{K}}$. We thus conclude that the parameter domain for CRM(\mathcal{K}) consists of the Cartesian product

$$\mathcal{D} = \Lambda_{\mathcal{K}} \times [0, 1]^Q \times [0, 1]^Q. \quad (4.38)$$

The outcome space is $[0, 1]^Q$, each coordinate of an outcome point being a possible probability for a correct answer to the corresponding item. The coordinate of a predicted point, given in Equation (4.36), may be rewritten as follows:

$$\tau(q) = (1 - \beta_q) \cdot \pi(\mathcal{K}_q) + \eta_q \cdot \pi(\mathcal{K}_{\bar{q}}). \quad (4.39)$$

Here is a particular example of the CRM.

Example 4.45 Consider the knowledge structure (Q, \mathcal{K}) with

$$Q = \{a, b, c\}, \quad (4.40)$$

and

$$\mathcal{K} = \{\emptyset, \{a\}, \{a, b\}, \{a, c\}, Q\}. \quad (4.41)$$

In CRM(\mathcal{K}), take item b ; we have for the b component of a predicted point

$$\begin{aligned} \tau(b) &= (1 - \beta_b) \cdot (\pi_{\{a,b\}} + \pi_Q) + \eta_b \cdot (\pi_{\emptyset} + \pi_{\{a\}} + \pi_{\{a,c\}}) \\ &= (1 - \beta_b) \cdot \pi(\mathcal{K}_b) + \eta_b \cdot \pi(\mathcal{K}_{\bar{q}}). \end{aligned}$$

The CRM makes an implicit apparition on page 270 of Doignon and Falmagne (1999), and is further investigated in Doignon (2017). We next define two particular submodels.

Definition 4.46 The CRM(\mathcal{K}) from Definition 4.44 is *fair* when the parameters η_q , for $q \in Q$, are set to 0. It is *straight* if the parameters β_q , for $q \in Q$, are also set to 0.

The coordinate w.r.t. q of a predicted point τ simplifies in a fair CRM to

$$\tau(q) = (1 - \beta_q) \cdot \pi(\mathcal{K}_q), \quad (4.42)$$

⁹ By definition, $\mathbb{R}^{\mathcal{K}}$ is formed by all functions from \mathcal{K} to \mathbb{R} ; it admits the basis consisting of all the vectors e_K for $K \in \mathcal{K}$, where e_K is the function mapping K to 1 and all other elements of \mathcal{K} to 0.

and further in a straight CRM to

$$\tau(q) = \pi(\mathcal{K}_q). \quad (4.43)$$

We first analyze the testability of straight CRMs. Given the knowledge structure (Q, \mathcal{K}) , the straight CRM(\mathcal{K}) is specified as $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ (because for each item q we have $\beta_q = \eta_q = 0$, there is no need to mention the latter parameters). Moreover, Equation (4.43) is also

$$\tau(q) = \sum \{\pi_K \mid K \in \mathcal{K}, q \in K\}. \quad (4.44)$$

Thus, the expression for the coordinate $\tau(q)$ is linear in the parameters π_K (even more: it is just a sum of some of the values π_K). We infer that the prediction function f is linear, or more adequately stated: f is the restriction to the simplex $\Lambda_{\mathcal{K}}$ of a linear function from the real linear space $\mathbb{R}^{\mathcal{K}}$ to the real linear space \mathbb{R}^Q . The said linear function is unique (because $\Lambda_{\mathcal{K}}$ linearly generates $\mathbb{R}^{\mathcal{K}}$), hence we denote it by the same letter f . It admits the description

$$f : \mathbb{R}^{\mathcal{K}} \rightarrow \mathbb{R}^Q : \pi \mapsto \tau = M_{\mathcal{K}} \cdot \pi, \quad (4.45)$$

where $M_{\mathcal{K}}$ is the *incidence matrix* of the structure (Q, \mathcal{K}) ; the latter matrix has rows indexed by items and columns by states; for $q \in Q$ and $K \in \mathcal{K}$, we have

$$(M_{\mathcal{K}})_{q,K} = \begin{cases} 0 & \text{if } q \notin K, \\ 1 & \text{if } q \in K. \end{cases} \quad (4.46)$$

Thus, $M_{\mathcal{K}}$ is the matrix representing the linear mapping f in the canonical basis of $\mathbb{R}^{\mathcal{K}}$ and \mathbb{R}^Q .

Example 4.47 For the straight CRM based on the knowledge structure (Q, \mathcal{K}) from Example 4.45, the prediction function is described by the following matrix equation:

$$\begin{pmatrix} \tau(a) \\ \tau(b) \\ \tau(c) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \pi_{\emptyset} \\ \pi_{\{a\}} \\ \pi_{\{a,b\}} \\ \pi_{\{a,c\}} \\ \pi_Q \end{pmatrix}. \quad (4.47)$$

Notice that the linear mapping $f : \mathbb{R}^{\mathcal{K}} \rightarrow \mathbb{R}^Q$ maps the vertices of the simplex $\Lambda_{\mathcal{K}}$ to some of the vertices of the cube $[0, 1]^Q$. Indeed, any vertex of $\Lambda_{\mathcal{K}}$ is a probability distribution on \mathcal{K} that concentrates its mass on some state K in \mathcal{K} ; as before, we denote this vertex by e_K . Equation (4.44) entails that the image of the vertex e_K of $\Lambda_{\mathcal{K}}$ is the vector $f(e_K) = \tau$ in $[0, 1]^Q$ with coordinate $\tau(q)$, for $q \in Q$, equal to 1 if $q \in K$, to 0 otherwise. In other words, $f(e_K)$ is in \mathbb{R}^Q the *characteristic vector* χ_K of K .

To summarize the previous paragraph, the mapping f is linear and it maps any vertex e_K of $\Lambda_{\mathcal{K}}$ to the vertex χ_K of the unit cube in \mathbb{R}^Q . Consequently, the prediction range of the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ is the convex hull in \mathbb{R}^Q of all

characteristic vectors χ_K , for K in \mathcal{K} . The prediction range is thus a 0/1-polytope (as defined in Subsection 4.3.1), which we call the *polytope of the knowledge structure* \mathcal{K} and denote as $P_{\mathcal{K}}$. Here is a statement of our previous, easy findings.

Proposition 4.48 *For a straight correct response model $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$, the prediction range $f(\mathcal{D})$ is the 0/1-polytope $P_{\mathcal{K}}$ whose vertices are the characteristic vectors of the states in \mathcal{K} .*

For a straight CRM based on the structure (Q, \mathcal{K}) , any linear description (by affine equations and inequalities) of the polytope $f(\mathcal{D})$ is thus a description of the prescription range $f(\mathcal{D})$. Here is an example of a straight CRM, which illustrates the criterion for testability that we next give in Proposition 4.50.

Example 4.49 Let $Q = \{a, b\}$ and $\mathcal{K} = \{\emptyset, \{a\}, \{b\}, Q\}$. Then the prediction function f of the straight CRM(\mathcal{K}) is surjective from the three-dimensional simplex $\Lambda_{\mathcal{K}}$ onto the full square $[0, 1]^Q$. The model is not testable.

Proposition 4.50 *The straight correct response model $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ is testable if and only if $\mathcal{K} \neq 2^Q$.*

Proof From Proposition 4.48, the prediction range of the model under consideration is the polytope $P_{\mathcal{K}}$ having as vertices the characteristic vectors of the subsets of Q that belong to \mathcal{K} . Now $P_{\mathcal{K}}$ differs from the unit cube $[0, 1]^{\mathcal{K}}$ exactly if some vertex of the cube is not a vertex of $P_{\mathcal{K}}$, in other words the characteristic vector of some subset N of Q is not a vertex of $P_{\mathcal{K}}$. This means $N \notin \mathcal{K}$. \square

We now investigate the quantitative testability of the straight CRM. Our crucial tool will be the rank of the linear mapping $f : \mathbb{R}^{\mathcal{K}} \rightarrow \mathbb{R}^Q$, that is, the rank of the incidence matrix $M_{\mathcal{K}}$ of (Q, \mathcal{K}) (the matrix was defined in Equation (4.46)).

Example 4.51 For the CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ from Example 4.49, the incidence matrix is the 0/1-matrix in

$$\begin{array}{c|cccc} & \emptyset & \{a\} & \{b\} & Q \\ \hline a & 0 & 1 & 0 & 1 \\ b & 0 & 0 & 1 & 1 \end{array} \quad (4.48)$$

The incidence matrix has rank 2, and so the linear mapping f has rank 2. Because $|Q| = 2$, the model is not quantitatively testable in view of the next proposition (anyway, we know from Example 4.49 that the prediction range coincides here with the outcome range).

Proposition 4.52 *Assume $|Q| > 1$. The straight correct response model $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ is quantitatively testable if and only if the rank of the incidence matrix $M_{\mathcal{K}}$ is strictly less than $|Q|$.*

Proof Here the prediction range is the polytope $P_{\mathcal{K}}$ lying in \mathbb{R}^Q . Because the origin is the vertex χ_{\emptyset} of $P_{\mathcal{K}}$, the affine subspace A generated by $P_{\mathcal{K}}$ coincides with the linear subspace generated by its vertices. Thus, the dimension of A equals

the rank of the matrix $M_{\mathcal{K}}$. Hence $P_{\mathcal{K}}$ has null measure if and only if $M_{\mathcal{K}}$ has rank less than the dimension $|Q|$ of the space \mathbb{R}^Q . \square

An elementary criterion for the testability of a straight CRM appears in Proposition 4.50. To the contrary, the criterion for quantitative testability in Proposition 4.52 looks rather technical (although it is implementable with any rank-computing algorithm). There is an intricate combinatorial rephrasing of quantitative testability which we skip here (see Doignon, 2017).

What about the non-straight case? Remember that the definition of the general CRM sets the value of $\tau(q)$ in Equation (4.39), that is

$$\tau(q) = (1 - \beta_q) \cdot \pi(\mathcal{K}_q) + \eta_q \cdot \pi(\mathcal{K}_{\bar{q}}). \quad (4.49)$$

If we impose $\pi_Q = 1$, we get $\tau(q) = 1 - \beta_q$. If we let β_q take all values in $[0, 1]$, the prediction range of the model is thus the whole unit cube $[0, 1]^Q$ (even if $\eta_q = 0$ for each item q). Thus, the fair CRM is not testable, and of course neither is the general CRM.

Thus, without restrictions on the β_q 's and η_q 's, the CRM predicts all possible values of the outcome probabilities. Of course, it does not make sense to let β_q and η_q vary in the whole interval $[0, 1]$. We will thus restrict their values to be small in some sense. A meaningful assumption is that they vary in a subinterval $[0, \alpha]$, where $0 < \alpha < 0.5$. By accordingly restricting the parameter domain and the prediction function in the CRM, we get the α -CRM.

Proposition 4.53 *The α -correct response model $(\Lambda_{\mathcal{K}} \times [0, \alpha]^Q \times [0, \alpha]^Q, f, [0, 1]^Q)$ is testable if and only if $\mathcal{K} \neq 2^Q$.*

Proof The prediction range of the α -CRM $(\Lambda_{\mathcal{K}} \times [0, \alpha]^Q \times [0, \alpha]^Q, f, [0, 1]^Q)$ always contains the prediction range of the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$. Thus if $\mathcal{K} = 2^Q$, neither model is testable (see Proposition 4.50 for the CRM).

Conversely, assume $\mathcal{K} \neq 2^Q$ and pick some subset N of Q not in \mathcal{K} . To show that our α -CRM is testable, we prove that the characteristic vector χ_N is not predicted by the model. Assume to the contrary that, for some vector (π, β, η) in $\Lambda_{\mathcal{K}} \times [0, \alpha]^Q \times [0, \alpha]^Q$, we have both

$$q \in N \implies 1 = (1 - \beta_q) \cdot \pi(\mathcal{K}_q) + \eta_q \cdot \pi(\mathcal{K}_{\bar{q}}), \quad (4.50)$$

$$r \in Q \setminus N \implies 0 = (1 - \beta_r) \cdot \pi(\mathcal{K}_r) + \eta_r \cdot \pi(\mathcal{K}_{\bar{r}}). \quad (4.51)$$

The quantities in the equality from (4.50) satisfy $1 - \beta_q \leq 1$, $\eta_q \leq \alpha < 1$ and $0 \leq \pi(\mathcal{K}_q) \leq 1$, and moreover $\pi(\mathcal{K}_q) + \pi(\mathcal{K}_{\bar{q}}) = 1$. The right-hand side is thus a convex combination of $1 - \beta_q$ and η_q . It can be equal to 1 only if $\pi(\mathcal{K}_q) = 1$. Similarly, we derive $\pi(\mathcal{K}_r) = 0$ from (4.51). For K in \mathcal{K} , the last equality implies $\pi_K = 0$ if $K \not\subseteq N$, and then the previous one implies $\pi_K = 0$ if $K \subset N$. We conclude that the mass of π must be concentrated on N , thus $N \in \mathcal{K}$, the required contradiction. \square

The α -CRM (and consequently the CRM) never satisfies quantitative testability, because its prediction range always contains the cube $[0, \alpha]^Q$.

4.7 The Correct Response Model: Identifiability

As early as 1987, M. Landy (private communication) noticed that in the straight case, different sets of values of the parameters can predict the same family of correct response probabilities. In other words, even in the straight case, a model CRM(\mathcal{K}) can be non-identifiable.

Example 4.54 Assume

$$Q \supseteq \{a, b, c, d\}, \quad \mathcal{K} \supseteq \{\{a, b\}, \{c, d\}, \{a, c\}, \{b, d\}\}, \quad (4.52)$$

and define the two probability distributions π_1 and π_2 on \mathcal{K} by setting

$$\pi_1(\{a, b\}) = \pi_1(\{c, d\}) = 0.5, \quad \pi_2(\{a, c\}) = \pi_2(\{b, d\}) = 0.5. \quad (4.53)$$

For any choice of values for the β_q 's and η_q 's, the model CRM(\mathcal{K}) gives the same predictions for either π_1 or π_2 : in both cases, the correct response probability equals $(1 - \beta_q)0.5 + \eta_q 0.5$ for any item q in $\{a, b, c, d\}$, and η_q for any item q elsewhere. Thus, the CRM(\mathcal{K}) is non-identifiable at some points of its parameter domain.

As we discussed after Definition 4.3, non-identifiability of a model may be perceived as a problem. In KST, Example 4.54 (and many other similar situations, see Corollary 4.56) shows that the independent observation of the answers of a learner does not always allow a unique probability distribution on the collection of states to be uncovered. Although in practical situations we may well go on with non-identifiability, we would like to design a general way to restore identifiability of the straight CRM without modifying the set of predicted points. A first idea would point to a careful modification of the knowledge structure \mathcal{K} at hand. However, no such modification can work because the prediction range determines the structure \mathcal{K} . To explain this, we note that the prediction range is the convex polytope $P_{\mathcal{K}}$, and so it determines the vertices of the polytope; and the vertices happen to be the characteristic vectors of the states forming \mathcal{K} .

There is a general way of restoring identifiability of any model without altering the output range of the prediction function: it consists in restricting the parameter domain in a careful way. As was explained after Definition 4.2, it suffices to keep only one parameter point in each preimage of the prediction function. Later in the present section, we show how to perform such a restriction of the parameter domain of the straight CRM – at least for a large collection of knowledge structures. Let us first state a criterion for identifiability (again in the straight case).

Proposition 4.55 *A straight CRM ($\Lambda_{\mathcal{K}}, f, [0, 1]^Q$) is identifiable if and only if the incidence matrix $M_{\mathcal{K}}$ has rank equal to $|\mathcal{K}| - 1$.*

Proof The rank of the incidence matrix is the rank of the linear mapping $f : \mathbb{R}^{\mathcal{K}} \rightarrow \mathbb{R}^{\mathcal{Q}}$. By a classical result from linear algebra, the rank of f equals the dimension of the linear space $\mathbb{R}^{\mathcal{K}}$ minus the dimension of the kernel $L = f^{-1}(o)$ of f . It suffices thus to prove that the model is identifiable if and only if the kernel L is of dimension 1. Notice first that $f(e_{\emptyset}) = o$ (where e_{\emptyset} designates the vertex of the simplex $\Lambda_{\mathcal{K}}$ formed by the probability distribution concentrated on the empty set \emptyset), so that the dimension of L is always at least 1.

If the model is not identifiable, two parameter points π and π' have the same image. Then the vector $\pi' - \pi$ is in the kernel L of f , and not a multiple of the vector π_{\emptyset} ; hence, L has dimension greater than 1.

Conversely, assume the kernel L has dimension greater than 1. Take any point π in the relative interior of the simplex $\Lambda_{\mathcal{K}}$. In $\mathbb{R}^{\mathcal{K}}$, the translation mapping the origin to π sends the linear subspace L to some affine subspace A . Because it is of dimension at least 2, A and the affine hyperplane spanned by the vertices of $\Lambda_{\mathcal{K}}$ (which have π in common) have at least an affine line in common. Because π is in the relative interior of $\Lambda_{\mathcal{K}}$, there is a second parameter point π' in $A \cap \Lambda_{\mathcal{K}}$. Then the vector $\pi' - \pi$ lies in the kernel L , and the two parameter points π and π' have the same image by f . So the model is not identifiable. \square

Corollary 4.56 *If the knowledge structure $(\mathcal{Q}, \mathcal{K})$ has at least two more states than items (that is, $|\mathcal{K}| \geq |\mathcal{Q}| + 2$), the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^{\mathcal{Q}})$ is not identifiable.*

Proof Proceeding by contraposition, suppose the model is identifiable. Then, by Proposition 4.55, the rank of the matrix $M_{\mathcal{K}}$ equals $|\mathcal{K}| - 1$. Notice that the column of $M_{\mathcal{K}}$ indexed by the empty state has only zero entries, so that after deletion of that column the resulting matrix still has rank $|\mathcal{K}| - 1$. Now the rank of the matrix has to be less or equal to the number of rows, which is here equal to $|\mathcal{Q}|$. Thus $|\mathcal{K}| \leq |\mathcal{Q}| + 1$. \square

Corollary 4.56 states that if the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^{\mathcal{Q}})$ is identifiable, then $|\mathcal{K}| \leq |\mathcal{Q}| + 1$. The converse does not hold: if we take $\mathcal{Q} = \{a, b, c, d\}$ and $\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \mathcal{Q}\}$, then the matrix $M_{\mathcal{K}}$ becomes

	\emptyset	$\{a\}$	$\{b\}$	$\{a, b\}$	\mathcal{Q}	
a	0	1	0	1	1	
b	0	0	1	1	1	
c	0	0	0	0	1	
d	0	0	0	0	1	

(4.54)

So $M_{\mathcal{K}}$ has rank 3, and by Proposition 4.55 the model is not identifiable.

We now introduce a way of restoring identifiability of a straight CRM which is not identifiable, under the assumption that the underlying knowledge structure is quasi-ordinal in the sense of Definition 4.31 (we will indicate that our method does not work for other structures). The method consists of restricting the parameter domain in a rather coherent way, and this will be achieved without modifying the prediction range.

Assume the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ is not identifiable, and moreover the structure \mathcal{K} is quasi-ordinal. By Theorem 4.33, the states of \mathcal{K} are exactly the beginning sets of a unique quasi-order \preceq on Q . In rough terms, we will replace the parameter domain, which is the simplex $\Lambda_{\mathcal{K}}$, with the union of some of its faces, and next replace the prediction function with its restriction to this union. The faces we take to form the union correspond to the maximal chains in the set \mathcal{K} ordered by inclusion. More precisely, if \mathcal{C} is any maximal chain in (\mathcal{K}, \subseteq) , we consider the face $F_{\mathcal{C}}$ of $\Lambda_{\mathcal{K}}$ whose vertices are the e_K , for $K \in \mathcal{C}$.

Any chain \mathcal{C} of subsets of Q bijectively correspond to a weak order on Q (see Birkhoff Theorem 4.33); let us denote this weak order by $\preceq_{\mathcal{C}}$. Moreover, the chain \mathcal{C} is contained in the quasi-ordinal structure \mathcal{K} if and only if the weak order $\preceq_{\mathcal{C}}$ is an extension of the quasi-order \preceq ; also, it is a maximal chain in \mathcal{K} exactly if $\preceq_{\mathcal{C}}$ is a tight extension of \preceq . To summarize, any maximal chain \mathcal{C} in (\mathcal{K}, \subseteq) corresponds to a tight extension of \preceq ; on the other hand, it also delivers the face $F_{\mathcal{C}}$ of the simplex $\Lambda_{\mathcal{K}}$. It happens that the region $f(F_{\mathcal{C}})$ (remember that f is the prediction function) is described by linear inequalities which “represent” the weak order $\preceq_{\mathcal{C}}$: for $x \in [0, 1]^Q$, we have $x \in f(F_{\mathcal{C}})$ exactly if $x_p \geq x_q$ holds for all p, q in Q such that $p \preceq_{\mathcal{C}} q$. Let us illustrate this on an example.

Example 4.57 Take a straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$ with $\cup \mathcal{K} = Q = \{a, b, c, d, e\}$ and a quasi-ordinal structure \mathcal{K} . Assume \mathcal{K} contains the chain \mathcal{C} of subsets

$$\emptyset \subset \{a\} \subset \{a, b, c\} \subset Q. \quad (4.55)$$

Suppose we set to 0 all values of $\pi(L)$ for L in $\mathcal{K} \setminus \mathcal{C}$, in other words we consider only the parameter points which are in the face $F_{\mathcal{C}}$ of the simplex $\Lambda_{\mathcal{K}}$. The action of the prediction function f on $F_{\mathcal{C}}$ is captured by the following matrix equation:

$$\begin{pmatrix} \tau(a) \\ \tau(b) \\ \tau(c) \\ \tau(d) \\ \tau(e) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \pi_{\emptyset} \\ \pi_{\{a\}} \\ \pi_{\{a,b,c\}} \\ \pi_Q \end{pmatrix}. \quad (4.56)$$

Thus, the predicted points τ which are in $f(F_{\mathcal{C}})$ all satisfy

$$\tau(a) \geq \tau(b) = \tau(c) \geq \tau(d) = \tau(e). \quad (4.57)$$

Notice the link with the weak order $\preceq_{\mathcal{C}}$ coming from the chain \mathcal{C} : we have $\tau(p) \geq \tau(q)$ exactly if $p \preceq_{\mathcal{C}} q$. Conversely, any point x in $[0, 1]^Q$ satisfying $x_a \geq x_b = x_c \geq x_d = x_e$ is obtained from a parameter point satisfying $\pi_K = 0$ for K in $\mathcal{K} \setminus \mathcal{C}$: it suffices to set

$$\pi_Q = x_e, \quad (4.58)$$

$$\pi_{\{a,b,c\}} = x_c - x_e, \quad (4.59)$$

$$\pi_{\{a\}} = x_a - x_c, \quad (4.60)$$

$$\pi_{\emptyset} = 1 - x_a. \quad (4.61)$$

Moreover, Equations (4.56) and (4.58)–(4.61) set those points x in a one-to-one correspondence with the parameter points π in F_C .

Example 4.57 suggests a way to ensure a one-to-one correspondence between the parameter points in the face F_C (some subset of the parameter domain of the CRM) and a subset of the points that this CRM predicts. Notice also how the structure of the chain of states is related to inequalities satisfied by the coordinates of the resulting predicted points. We now describe the general method.

Assume (Q, \mathcal{K}) is a quasi-ordinal space, and consider the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$. For any maximal chain \mathcal{C} of states in \mathcal{K} , take the face $F_{\mathcal{C}}$ of the simplex $\Lambda_{\mathcal{K}}$ with vertices e_K for K in \mathcal{C} . Then set $\mathcal{E}_{\mathcal{K}} = \cup_{\mathcal{C}} F_{\mathcal{C}}$ (hence $\mathcal{E}_{\mathcal{K}}$ is a subset of $\Lambda_{\mathcal{K}}$), and let g be the restriction of f to $\mathcal{E}_{\mathcal{K}}$. We obtain a new probabilistic model $(\mathcal{E}_{\mathcal{K}}, g, [0, 1]^Q)$. The next statement asserts that the new model has the required properties.

Proposition 4.58 *For any quasi-ordinal space (Q, \mathcal{K}) , the probabilistic model $(\mathcal{E}_{\mathcal{K}}, g, [0, 1]^Q)$ has the same prediction range as the straight CRM $(\Lambda_{\mathcal{K}}, f, [0, 1]^Q)$. Moreover, it is identifiable.*

Proof Again, let \mathcal{K} consist of all the beginning sets of the quasi-order \preccurlyeq on Q . Because g is the restriction of f to the subdomain $\mathcal{E}_{\mathcal{K}}$ of $\Lambda_{\mathcal{K}}$, we clearly have $g(\mathcal{E}_{\mathcal{K}}) \subseteq f(\Lambda_{\mathcal{K}})$. To prove the reverse inclusion, let x be any point in $f(\Lambda_{\mathcal{K}})$. We form a weak order \preceq on Q by letting $p \preceq q$ exactly if $x_p \geq x_q$, for p, q in Q . Then \preceq is an extension of the quasi-order \preccurlyeq . Indeed, if $p \preccurlyeq q$ for some p, q in Q , then $\mathcal{K}_p \supseteq \mathcal{K}_q$, and so by the definition of f we have $x_p \geq x_q$, thus $p \preceq q$. It follows that \preceq equals \preceq_B for some chain B in \mathcal{K} , that is: $x_p \geq x_q$ if and only if $B_p \supseteq B_q$. Now let \mathcal{C} be any maximal chain in \mathcal{K} with $B \subseteq \mathcal{C}$. Thus, the face $F_{\mathcal{C}}$ is among the faces of which we form the union to get $\mathcal{E}_{\mathcal{K}}$. It suffices now to check that some point π of $F_{\mathcal{C}}$, that is, some probability distribution on \mathcal{K} which is null on $\mathcal{K} \setminus \mathcal{C}$, satisfies $g(\pi) = f(\pi) = x$. If we write down the system in the unknowns $\pi(C)$, for $C \in \mathcal{C}$, sorting the equations in such a way that the independent terms $x(p)$ are in non-increasing order,

$$x(p) = \sum_{C \in \mathcal{C}, p \in C} \pi(C), \quad \text{for } p \in Q, \tag{4.62}$$

and add $\sum_{C \in \mathcal{C}} \pi(C) = 1$ as the very first equation, we see that the coefficients of the system are all 0 or 1, with the 1 above the 0 (the separation is staircase-like). Consequently, for any given independent terms $x(p)$, there is always a solution π with $\pi(C) \geq 0$ for C in \mathcal{C} and $\pi(L) = 0$ for L in $\mathcal{K} \setminus \mathcal{C}$.

To prove that the new model $(\mathcal{E}_{\mathcal{K}}, g, [0, 1]^Q)$ is identifiable, we again take any x in $g(\mathcal{E}_{\mathcal{K}})$, that is, in $f(\Lambda_{\mathcal{K}})$, and prove that there is only one point π in $\mathcal{E}_{\mathcal{K}}$ such that $x = g(\pi)$. We use the chains B and \mathcal{C} as in the previous paragraph. It is easily checked that the solution π we found satisfies $\pi(C) = 0$ for $C \in \mathcal{C} \setminus B$, and its values $\pi(B)$, for $B \in B$, are unequivocally determined. Hence, whatever the selection of the maximal chain \mathcal{C} , we always get, for the predicted point x given in $g(\mathcal{E}_{\mathcal{K}})$, the same parameter point π . \square

Why does our method of reducing the parameter domain to obtain an identifiable model work only under the assumption that the structure be quasi-ordinal? The essential reason is that the quasi-order inferred from the coordinates of the predicted point under scrutiny is linked to some maximal chain of states. As a matter of fact, there is a geometrical result underlying our construction. A proposition in Stanley (1986) can be rephrased in our terms as follows. Assume (Q, \mathcal{K}) is a finite ordinal space, with its states consisting of the beginning sets of the partial order \leq on Q . Consider the straight CRM $(\Delta_{\mathcal{K}}, f, [0, 1]^Q)$. The prediction range $f(\Delta_{\mathcal{K}})$ is then the union of all the simplices of the following special type: for each linear extension L of \leq , take the simplex S_L formed by all points x of $[0, 1]^Q$ such that $x_p \geq x_q$, for $p L q$. Then $f(\Delta_{\mathcal{K}})$ is the union of all these simplices. Moreover, the intersection of any two of these simplices forms a face of each of them. Also, any vertex of such a simplex is a vertex of $f(\Delta_{\mathcal{K}})$. The latter three conditions mean that $f(\Delta_{\mathcal{K}})$ is *triangulated* by the simplices S_L , for L varying in the linear extensions of \leq . Proposition 4.58 relies on a similar triangulation in the slightly more general context of a quasi-ordinal space. It is shown in Doignon (2017) that only for quasi-ordinal spaces does the approach followed above work.

We now briefly turn to a more general (not necessarily straight) correct response model, namely the α -CRM (see Proposition 4.53). The α -CRM is never identifiable, as one can easily show by varying the values of the β_q 's and η_q 's. In this setting we do not know of any natural way to restore identifiability.

4.8 The Basic Local Independence Model

As we saw on the toy example in Section 4.1, the basic local independence model (BLIM) predicts the probability of each response pattern R ; here the pattern R , a subset of the domain Q , contains all the items that would be answered correctly. The set of all response patterns is denoted by \mathcal{R} ; thus $\mathcal{R} = 2^Q$. The fundamental assumption in the BLIM is local stochastic independence, which means that the responses across items are stochastically independent given the knowledge state (Falmagne & Doignon, 1988).

Definition 4.59 Given a knowledge structure (Q, \mathcal{K}) , the *basic local independence model* (BLIM) based on \mathcal{K} , denoted as $(Q, \mathcal{K}, \pi, \beta, \eta)$ or in short $\text{BLIM}(\mathcal{K})$, has parameters π_K , β_q , η_q which, respectively, hold the probability values of a state K , a careless error and a lucky guess in an answer to item q . It predicts that the items correctly answered form the pattern R in \mathcal{R} with probability

$$\rho(R) = \sum_{K \in \mathcal{K}} \rho(R | K) \cdot \pi_K, \quad (4.63)$$

where for all $K \in \mathcal{K}$

$$\rho(R \mid K) = \left(\prod_{q \in K \setminus R} \beta_q \right) \cdot \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \cdot \left(\prod_{q \in R \setminus K} \eta_q \right) \cdot \left(\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right). \quad (4.64)$$

There is a matrix formulation for Equation 4.63, as in the next example.

Example 4.60 Assume $Q = \{a, b\}$ and $\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$. Then the following equation characterizes the predictive function of $\text{BLIM}(\mathcal{K})$.

$$\begin{aligned} & \begin{pmatrix} \rho(\emptyset) \\ \rho(\{a\}) \\ \rho(\{b\}) \\ \rho(\{a, b\}) \end{pmatrix} \\ &= \begin{pmatrix} (1 - \eta_a)(1 - \eta_b) & \beta_a(1 - \eta_b) & (1 - \eta_a)\beta_b & \beta_a\beta_b \\ \eta_a(1 - \eta_b) & (1 - \beta_a)(1 - \eta_b) & \eta_a\beta_b & (1 - \beta_a)\beta_b \\ (1 - \eta_a)\eta_b & \beta_a\eta_b & (1 - \eta_a)(1 - \beta_b) & \beta_a(1 - \beta_b) \\ \eta_a\eta_b & (1 - \beta_a)\eta_b & \eta_a(1 - \beta_b) & (1 - \beta_a)(1 - \beta_b) \end{pmatrix} \cdot \begin{pmatrix} \pi_\emptyset \\ \pi_{\{a\}} \\ \pi_{\{b\}} \\ \pi_{\{a, b\}} \end{pmatrix}. \end{aligned} \quad (4.65)$$

In the fair case, the equation simplifies to

$$\begin{pmatrix} \rho(\emptyset) \\ \rho(\{a\}) \\ \rho(\{b\}) \\ \rho(\{a, b\}) \end{pmatrix} = \begin{pmatrix} 1 & \beta_a & \beta_b & \beta_a\beta_b \\ 0 & 1 - \beta_a & 0 & (1 - \beta_a)\beta_b \\ 0 & 0 & 1 - \beta_b & \beta_a(1 - \beta_b) \\ 0 & 0 & 0 & (1 - \beta_a)(1 - \beta_b) \end{pmatrix} \cdot \begin{pmatrix} \pi_\emptyset \\ \pi_{\{a\}} \\ \pi_{\{b\}} \\ \pi_{\{a, b\}} \end{pmatrix}. \quad (4.66)$$

The matrix in Equation (4.66) is upper triangular. In the fair case with $\mathcal{K} = 2^Q$, such a form can always be obtained (it suffices to list the patterns in such a way that if $R \subset R'$, then R is listed before R' , and similarly for the states).

Characterizing the general $\text{BLIM}(Q, \mathcal{K}, \pi, \beta, \eta)$ as a probabilistic model amounts to specifying its parameter domain and outcome space as well as its prediction function. We do this in the next three subsections.

4.8.1 Parameter Domain

The parameter domain consists of the vectors $\theta = (\pi, \beta, \eta)$, where $\beta = (\beta_q)_{q \in Q}$ and $\eta = (\eta_q)_{q \in Q}$ denote the vectors of item-specific careless error and lucky guessing probabilities. In order to get non-redundant parameters, all but one of the state probabilities π_K , for K in \mathcal{K} , are included in the corresponding parameter vector π . The natural choice for excluding a state from \mathcal{K} is among those subsets of the domain that are definitely in \mathcal{K} , which are the naïve state \emptyset and the state Q of full mastery. The following refers to $\mathcal{K}^* = \mathcal{K} \setminus \{Q\}$. Thus, let $\pi = (\pi_K)_{K \in \mathcal{K}^*}$ denote the parameter vector of non-redundant state probabilities π_K , $K \in \mathcal{K}^*$.

In general, the parameter vectors $\theta = (\pi, \beta, \eta)$ consist of $m = (|\mathcal{K}| - 1) + 2 \cdot |Q|$ components, m being the number of non-redundant parameters. The subsequently

imposed constraints either follow from basic assumptions of KST or are convenient for characterizing the testability and identifiability of the BLIM. In any case, these reasonable constraints restrict the parameter space to a proper subset of the m -dimensional unit hypercube. Here they are:

$$0 < \pi_K, \quad K \in \mathcal{K}^*, \quad (\text{C1})$$

$$\sum_{L \in \mathcal{K}^*} \pi_L < 1, \quad (\text{C2})$$

$$0 < \beta_q < 1, \quad q \in Q, \quad (\text{C3})$$

$$0 < \eta_q < 1, \quad q \in Q, \quad (\text{C4})$$

$$\beta_q + \eta_q < 1, \quad q \in Q. \quad (\text{C5})$$

The inequality in (C5) means nothing else but that a correct response is more likely if the item is mastered than if it is not mastered, which is at the very heart of the idea of a knowledge state.

Consequently,

$$\mathcal{D} = \{\theta \in \mathbb{R}^m \mid \theta = (\pi, \beta, \eta) \text{ satisfies (C1) -- (C5)}\} \quad (4.67)$$

is an open, convex subset of \mathbb{R}^m , and it is thus connected.

We impose a strict inequality in (C5) in order to discard a BLIM in which the condition $\beta_q = \eta_q = 0.5$ holds for all items q . Such a special BLIM satisfies

$$\rho(R) = \sum_{K \in \mathcal{K}} \rho(R \mid K) \pi_K = 0.5^{|Q|} \sum_{K \in \mathcal{K}} \pi_K = 0.5^{|Q|}, \quad (4.68)$$

and thus predicts a uniform probability distribution on the response patterns, independent of the underlying knowledge structure \mathcal{K} . It thus imposes no restrictions at all on the parameters π_K for $K \in \mathcal{K}^*$, which means that any BLIM on this extended parameter space is not identifiable.

4.8.2 Outcome Space

The outcome space \mathcal{O} may be conceived as the collection of all probability distributions on the set of response patterns \mathcal{R} , which is identified with the power set of the domain Q . Confining consideration to $\mathcal{R}^* = \mathcal{R} \setminus \{Q\}$ for ensuring non-redundancy, we get $n = |\mathcal{R}^*| = 2^{|Q|} - 1$ non-redundant observables. This leads to defining \mathcal{O} as the set of all $(\phi_R)_{R \in \mathcal{R}^*}$ in $[0, 1]^n$ satisfying $\sum_{R \in \mathcal{R}^*} \phi_R \leq 1$.

4.8.3 Prediction Function

The prediction function $f: \mathcal{D} \rightarrow \mathcal{O}$ associates to each parameter vector θ in \mathcal{D} a probability distribution $f(\theta)$ on \mathcal{R}^* , with

$$f(\theta) = (f_R(\theta))_{R \in \mathcal{R}^*} = (\rho(R))_{R \in \mathcal{R}^*}. \quad (4.69)$$

In terms of the non-redundant parameters $\theta = (\pi, \beta, \eta)$ we have

$$\begin{aligned} \rho(R) &= \\ &\sum_{K \in \mathcal{K}^*} \left(\prod_{q \in K \setminus R} \beta_q \right) \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \left(\prod_{q \in R \setminus K} \eta_q \right) \left(\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right) \pi_K \\ &\quad + \left(\prod_{q \notin R} \beta_q \right) \left(\prod_{q \in R} (1 - \beta_q) \right) \left(1 - \sum_{L \in \mathcal{K}^*} \pi_L \right), \end{aligned} \quad (4.70)$$

where the last term of the sum refers to the case $K = Q$. Notice that the expressions $f_R(\theta) = \rho(R)$ predicting the probability of the response pattern $R \in \mathcal{R}^*$ are sums of products of parameters. This implies that the prediction function f is polynomial, and thus analytic.

The questions of testability and identifiability (in the sense of Definitions 4.2 and 4.3) are more involved in the BLIM setting than in the CRM setting. The reason lies in the non-linearity of Equation (4.70). Our analysis of the straight CRM relied on Equation (4.39), which is linear as we explained earlier. Notice that the straight BLIM is not that interesting, because it sets

$$\rho(R) = \begin{cases} 0 & \text{if } R \notin \mathcal{K}, \\ \pi_K & \text{if } R = K \in \mathcal{K}. \end{cases} \quad (4.71)$$

Consequently, the straight BLIM is testable if and only if $\mathcal{K} \neq 2^Q$, and it is always identifiable. The sequel considers testability and identifiability of the general BLIM, including the fair case.

4.9 The Basic Local Independence Model: Testability

Notice that assuming state probabilities as well as response error probabilities to be non-zero implies that the predicted probabilities for all the response patterns are non-zero too. So the image $f(\mathcal{D})$ is a proper subset of \mathcal{O} . This means that any BLIM $(Q, \mathcal{K}, \beta, \eta, \pi)$ with its prediction function defined on the parameter domain \mathcal{D} is testable in the sense of Definition 4.2. This result is, of course, mainly due to the parameter constraints that were imposed in (C1)–(C5). We shall not delve more deeply into testability, but turn to the more interesting question of whether the BLIM is quantitatively testable in the sense of Definition 4.6. To be able to answer this question by drawing upon Proposition 4.20, the Jacobian matrix of the prediction function has to be determined. Lemma 4.61 characterizes the partial derivatives of f with respect to the parameters, which constitute its columns. For notational convenience define

$$\mu_q^R = \begin{cases} -\frac{1}{1-\beta_q} & \text{if } q \in R \\ \frac{1}{\beta_q} & \text{if } q \notin R \end{cases} \quad \sigma_q^R = \begin{cases} \frac{1}{\eta_q} & \text{if } q \in R \\ -\frac{1}{1-\eta_q} & \text{if } q \notin R \end{cases} \quad (4.72)$$

for all $R \in \mathcal{R}^*$. Moreover, as in Definition 4.25, let \mathcal{K}_q denote the set of all knowledge states in \mathcal{K} that contain item $q \in Q$, and let $\mathcal{K}_{\bar{q}} = \mathcal{K} \setminus \mathcal{K}_q$.

Lemma 4.61 *Consider a knowledge structure \mathcal{K} on the domain Q . The partial derivatives of the prediction function f of $\text{BLIM}(\mathcal{K}) = (Q, \mathcal{K}, \pi, \beta, \eta)$ are as follows for $R \in \mathcal{R}^*$ and, first, $K \in \mathcal{K}^*$,*

$$\frac{\partial f_R}{\partial \pi_K}(\theta) = \rho(R | K) - \rho(R | Q) \quad (4.73)$$

next for $q \in Q$,

$$\frac{\partial f_R}{\partial \beta_q}(\theta) = \mu_q^R \cdot \sum_{K \in \mathcal{K}_q} \rho(R | K) \cdot \pi_K \quad (4.74)$$

and

$$\frac{\partial f_R}{\partial \eta_q}(\theta) = \sigma_q^R \cdot \sum_{K \in \mathcal{K}_{\bar{q}}} \rho(R | K) \cdot \pi_K. \quad (4.75)$$

With this explicit characterization of the Jacobian matrix at hand, it is easily seen that quantitative testability of the BLIM depends in particular on the cardinality of the knowledge domain Q (we use Table 4.1 to draw conclusions). For $|Q| = 1$, the number $m = 3$ of non-redundant parameters is greater than the number $n = 1$ of non-redundant observables, and the maximum rank r of the Jacobian matrix of the prediction function will not fall below n ; so the model is not quantitatively testable. The same conclusion holds when $|Q| = 2$ (with $5 \leq m \leq 7$, depending on the knowledge structure \mathcal{K} , and $n = 3$). For $|Q| = 3$, there are 64 different knowledge structures, and we have $7 \leq m \leq 13$ and $n = 7$. For each of the (general) BLIMs defined on these knowledge structures, the maximum rank of the Jacobian matrix turns out to be $r = 7$ (Heller, 2017, table 1). We conclude that for $|Q| \leq 3$, the BLIM is never quantitatively testable. This picture changes if larger domains are considered.

Example 4.62 Let $|Q| = 4$, and consider $\mathcal{K} = \{\emptyset, Q\}$. Then we have $m = 9$ and $n = 15$, and the Jacobian matrix has full rank $r = 9$. So by Proposition 4.20, the corresponding BLIM is quantitatively testable.

Whenever the knowledge domain Q is large enough (i.e., $|Q| \geq 3$ for the fair BLIM, and $|Q| \geq 4$ for the general BLIM), the maximum rank of the Jacobian matrix of a BLIM (and thus its quantitative testability and local identifiability) depends upon combinatorial properties of the underlying knowledge structure. This is explored further in the subsequent section.

4.10 The Basic Local Independence Model: Identifiability

First of all, the inequality $m \leq n$ is a necessary condition for identifiability of the BLIM: it assures that the number m of non-redundant parameters does not exceed the number n of non-redundant observables (see Corollary 4.24). The

inequality is equivalent to $|\mathcal{K}| + 2 \cdot |\mathcal{Q}| \leq 2^{|\mathcal{Q}|}$. Because $2 \leq |\mathcal{K}|$ by definition, we derive that for $|\mathcal{Q}| < 3$, any fully parametrized BLIM cannot be identifiable.

There are interesting results on the general BLIM and arbitrary \mathcal{Q} which are due, in particular, to Spoto *et al.* (2012, 2013); Stefanutti *et al.* (2012), and Heller (2017). We report here only some of the findings. The first one relies on the combinatorial concept of backward-gradedness (Definition 4.38). A similar result (dual, in a technical sense) is available for forward-graded structures (Heller, 2017; Spoto *et al.*, 2012, 2013). The subsequent discussion treats the backward-graded case.

In the fair case, given a structure which is backward-graded in an item, Spoto *et al.* (2012) explicitly proved that the prediction function is not injective. Spoto *et al.* (2013, theorem 2) generalized the result by showing that the model in the general case is identifiable at no parameter point, and thus also at no predicted point (in the sense of Definition 4.3). The following is an adaptation of the latter result to respect the constraints (C1)–(C5) defining the parameter domain.

Proposition 4.63 *Let $(\mathcal{Q}, \mathcal{K}, \pi, \beta, \eta)$ be a BLIM, with $(\mathcal{Q}, \mathcal{K})$ backward-graded in q , and $\mathcal{K}_{-q} = \{K \setminus \{q\} \mid K \in \mathcal{K}_q\}$. Consider*

$$u = \min_{K \in \mathcal{K}_q} \left\{ \frac{\beta_q \pi_K + (1 - \eta_q) \pi_{K \setminus \{q\}}}{\pi_K + \pi_{K \setminus \{q\}}} \right\}.$$

Given a parameter vector $\theta = (\pi, \beta, \eta)$ in \mathcal{D} , define $\theta' = (\pi', \beta', \eta')$ by selecting an arbitrary value β'_q in the open interval $]0, u[$, and by setting $\eta'_q = \eta_q$. Otherwise, for all $p \neq q$ let

$$\eta'_p = \eta_p; \quad (4.76)$$

$$\beta'_p = \beta_p. \quad (4.77)$$

For $K \in \mathcal{K}_q$ define

$$\pi'_K = \frac{1 - \beta_q - \eta_q}{1 - \beta'_q - \eta_q} \pi_K, \quad (4.78)$$

$$\pi'_{K \setminus \{q\}} = \pi_{K \setminus \{q\}} + \frac{\beta_q - \beta'_q}{1 - \beta'_q - \eta_q} \pi_K; \quad (4.79)$$

and for $K \notin (\mathcal{K}_q \cup \mathcal{K}_{-q})$ let

$$\pi'_K = \pi_K. \quad (4.80)$$

Then $f(\theta) = f(\theta')$ and the BLIM is not identifiable on \mathcal{D} .

Proof Notice that $\theta' = (\pi', \beta', \eta')$ lies in \mathcal{D} . While constraint (C4) holds for η' by definition, it is easily seen that $\pi'_K + \pi'_{K \setminus \{q\}} = \pi_K + \pi_{K \setminus \{q\}}$ for all $K \in \mathcal{K}_q$, so that π' meets (C2). Moreover, drawing β'_q from the open interval $]0, u[$ guarantees that constraints (C1), (C3), and (C5) hold for θ' . In order to show that $f(\theta') = f(\theta)$, consider

$$f_R(\theta) = \sum_{K \in \mathcal{K}_q} \rho(R \mid K) \pi_K + \sum_{K \in \mathcal{K}_{-q}} \rho(R \mid K) \pi_K + \sum_{K \in \mathcal{K} \setminus (\mathcal{K}_q \cup \mathcal{K}_{-q})} \rho(R \mid K) \pi_K \quad (4.81)$$

for all $R \in 2^{\mathcal{Q}}$. Let us first turn to the last term. Due to (4.80) and $\eta'_q = \eta_q$ we have

$$\sum_{K \in \mathcal{K} \setminus (\mathcal{K}_q \cup \mathcal{K}_{-q})} \rho'(R \mid K) \pi'_K = \sum_{K \in \mathcal{K} \setminus (\mathcal{K}_q \cup \mathcal{K}_{-q})} \rho(R \mid K) \pi_K.$$

So, we may focus on the two remaining terms, and $f_R(\theta') = f_R(\theta)$ follows if

$$\rho'(R \mid K) \pi'_K + \rho'(R \mid K \setminus \{q\}) \pi'_{K \setminus \{q\}} = \rho(R \mid K) \pi_K + \rho(R \mid K \setminus \{q\}) \pi_{K \setminus \{q\}} \quad (4.82)$$

for all $K \in \mathcal{K}_q$. Now, let $q \notin R$. Then for each $K \in \mathcal{K}_q$ we have $\rho'(R \mid K) = C\beta'_q$, $\rho'(R \mid K \setminus \{q\}) = C(1 - \eta'_q)$, $\rho(R \mid K) = C\beta_q$, and $\rho(R \mid K \setminus \{q\}) = C(1 - \eta_q)$ with some $C \in]0, 1[$. Thus, due to (4.76) and (4.77) Equation (4.82) becomes

$$C\beta'_q \pi'_K + C(1 - \eta'_q) \pi'_{K \setminus \{q\}} = C\beta_q \pi_K + C(1 - \eta_q) \pi_{K \setminus \{q\}}. \quad (4.83)$$

Cancelling C and acknowledging that $\eta'_q = \eta_q$ provides

$$\beta'_q \pi'_K + (1 - \eta_q) \pi'_{K \setminus \{q\}} = \beta_q \pi_K + (1 - \eta_q) \pi_{K \setminus \{q\}} \quad (4.84)$$

for all $K \in \mathcal{K}_q$. Plugging in (4.78) and (4.79) into the left-hand side, a little algebra shows that it equals the right-hand side.

For $q \in R$, by essentially the same argument as above one obtains

$$(1 - \beta'_q) \pi'_K + \eta_q \pi'_{K \setminus \{q\}} = (1 - \beta_q) \pi_K + \eta_q \pi_{K \setminus \{q\}} \quad (4.85)$$

for all $K \in \mathcal{K}_q$. Again, plugging in (4.78) and (4.79) into the left-hand side establishes (4.85). \square

Proposition 4.63 shows that, if the knowledge structure is backward-graded in some item q , then for any point θ in the parameter domain \mathcal{D} there is another point θ' in \mathcal{D} such that $f(\theta) = f(\theta')$. The transformation of θ into θ' relies on a trade-off among the parameter values. For instance, it is always possible to increase the careless error probability β_q of the item q by some amount to β'_q , which, through Equations (4.78) and (4.79), can be compensated by increasing the probabilities of all those states $K \in \mathcal{K}_q$ that contain q by a corresponding amount, and by decreasing that of the states $K \setminus \{q\}$ accordingly.

For a (general) BLIM based on a knowledge structure \mathcal{K} which is backward-graded in item q , it is not a surprise then that the Jacobian matrix of $\text{BLIM}(\mathcal{K})$ is rank-deficient at any parameter point. This was first shown by Heller (2017, proposition 1), who exhibited the following linear dependence among the columns of the Jacobian matrix (where again $\mathcal{K}_{-q} = \{K \setminus \{q\} \mid K \in \mathcal{K}_q\}$): for all $R \in \mathcal{R}^*$,

$$(1 - \beta_q - \eta_q) \cdot \frac{\partial f_R}{\partial \beta_q}(\theta) - \sum_{K \in \mathcal{K}_{-q}} \pi_{K \cup \{q\}} \cdot \frac{\partial f_R}{\partial \pi_K}(\theta) + \sum_{K \in \mathcal{K}_q^*} \pi_K \cdot \frac{\partial f_R}{\partial \pi_K}(\theta) = 0. \quad (4.86)$$

Some consequences of this result are now illustrated by taking up Example 4.39.

Example 4.64 Consider the fair BLIM on the knowledge structure $\mathcal{K} = \{\emptyset, \{a, b\}, Q\}$ on $Q = \{a, b, c\}$ with the parameter vector

$$\theta = (\pi_\emptyset, \pi_{ab}, \beta_a, \beta_b, \beta_c). \quad (4.87)$$

The prediction function f of $\text{BLIM}(\mathcal{K})$ maps the parameter point $\theta = (\pi, \beta)$ to the predicted point¹⁰

$$\begin{pmatrix} \phi_\emptyset(\theta) \\ \phi_a(\theta) \\ \phi_b(\theta) \\ \phi_c(\theta) \\ \phi_{ab}(\theta) \\ \phi_{ac}(\theta) \\ \phi_{bc}(\theta) \end{pmatrix} = \begin{pmatrix} \pi_\emptyset + \beta_a \beta_b \pi_{ab} + \beta_a \beta_b \beta_c (1 - \pi_\emptyset - \pi_{ab}) \\ (1 - \beta_a) \beta_b \pi_{ab} + (1 - \beta_a) \beta_b \beta_c (1 - \pi_\emptyset - \pi_{ab}) \\ \beta_a (1 - \beta_b) \pi_{ab} + \beta_a (1 - \beta_b) \beta_c (1 - \pi_\emptyset - \pi_{ab}) \\ \beta_a \beta_b (1 - \beta_c) (1 - \pi_\emptyset - \pi_{ab}) \\ (1 - \beta_a) (1 - \beta_b) \pi_{ab} + (1 - \beta_a) (1 - \beta_b) \beta_c (1 - \pi_\emptyset - \pi_{ab}) \\ (1 - \beta_a) \beta_b (1 - \beta_c) (1 - \pi_\emptyset - \pi_{ab}) \\ \beta_a (1 - \beta_b) (1 - \beta_c) (1 - \pi_\emptyset - \pi_{ab}) \end{pmatrix}. \quad (4.88)$$

The Jacobian matrix J_f of the prediction function f turns out to be rank-deficient at all points, with everywhere in the parameter domain $\text{rk}(J_f) = 4$ being smaller than the number of parameters $m = 5$. The vector

$$v(\theta) = \left(0, 1, 0, 0, -\frac{1 - \beta_c}{1 - \pi_\emptyset - \pi_{ab}} \right) \quad (4.89)$$

generates the null space of the Jacobian matrix.

After picking a parameter point θ_0 , we form the Cauchy problem (a differential equation on the unknown function $t \mapsto \theta(t)$, where t is a real argument, together with an initial condition)

$$\begin{cases} \frac{d\theta}{dt}(t) = v(\theta), \\ \theta(0) = \theta_0. \end{cases} \quad (4.90)$$

Because $v(\theta)$ depends continuously on $\theta = (\pi_\emptyset, \pi_{ab}, \beta_a, \beta_b, \beta_c)$, our Cauchy problem admits a (locally unique) solution $\theta = \theta(t)$ with t varying in some open interval containing 0 (for the existence of a solution to a Cauchy problem, see for instance Lang, 1997, chapter XIX, or Dieudonné, 1960a, chapter 10). Then $t \mapsto \theta(t)$ parameterizes a trajectory or (piece of a) curve Γ in \mathcal{D} , which passes through θ_0 . The prediction function f remains constant on Γ ; in other words, the distribution $f(\theta)$ on the set \mathcal{R} of response patterns is the same for all parameter points θ on Γ . The zero entries in the vector $v(\theta)$ indicate constancy of the respective components of $f(\theta)$, while there is a trade-off between parameters π_{ab} and β_c . With $\theta_0 = (\pi_\emptyset^0, \pi_{ab}^0, \beta_a^0, \beta_b^0, \beta_c^0)$, the curve Γ is parameterized by $\theta(t) = (\pi_\emptyset^0, \pi_{ab}(t), \beta_a^0, \beta_b^0, \beta_c(t))$.

¹⁰ Note that we write point coordinates and vector components either in a row matrix or in a column matrix, depending on the need.

The functions β_c and π_{ab} are solutions of the two Cauchy problems

$$\begin{cases} \frac{d\pi_{ab}(t)}{dt} = 1, \\ \pi_{ab}(0) = \pi_{ab}^0, \end{cases} \quad \begin{cases} \frac{d\beta_c(t)}{dt} = -\frac{1 - \beta_c(t)}{1 - \pi_\varnothing - \pi_{ab}(t)}, \\ \beta_c(0) = \beta_c^0. \end{cases} \quad (4.91)$$

Finally, the curve Γ is parameterized by

$$\theta(t) = \left(\pi_\varnothing^0, \pi_{ab}^0 + t, \beta_a^0, \beta_b^0, \frac{\beta_c^0(1 - \pi_\varnothing^0 - \pi_{ab}^0) - t}{1 - \pi_\varnothing^0 - \pi_{ab}^0 - t} \right) \quad (4.92)$$

with t varying in the open interval $]-\pi_{ab}^0, \beta_c^0(1 - \pi_\varnothing^0 - \pi_{ab}^0)[$.

The arrows in Figure 4.7 illustrate the vector field¹¹ that the differential equation from (4.90) induces in the (β_c, π_{ab}) -plane of the parameter domain; here, the initial value π_\varnothing^0 was set to 0.4 while β_a^0 and β_b^0 are arbitrary. The solutions of the Cauchy problem parameterize the curves shown in the figure, which are all exemplars of Γ . It is easily checked that, as announced, the prediction function is

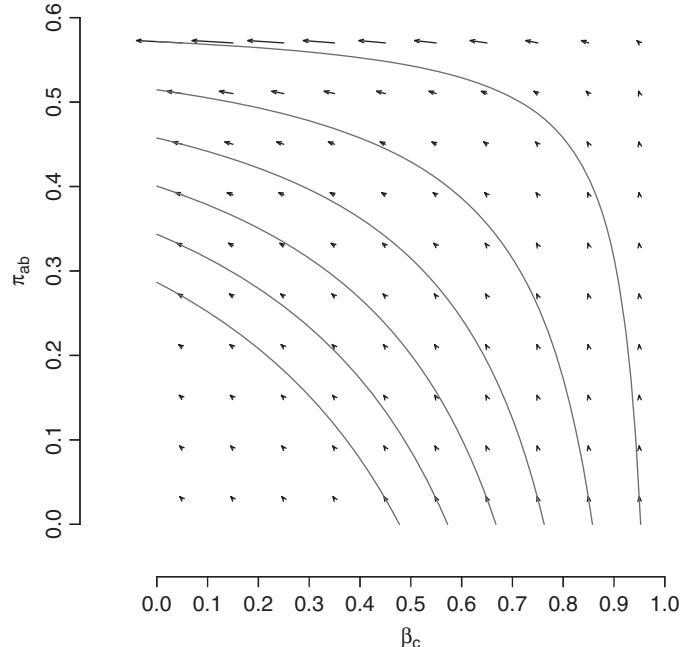


Figure 4.7 Vector field in the (β_c, π_{ab}) -plane with $\pi_\varnothing = 0.4$, and arbitrary parameter values β_a, β_b for the fair BLIM on knowledge structure $\mathcal{K}^{(3)}$. The curves represent particular solutions to the corresponding differential equations and characterize the trade-off between parameters β_c and π_{ab} .

¹¹ Here, we use the term “vector” with a different meaning: such a vector has an origin which in most of the cases differs from the zero point.

constant along these curves. As an aside, it may be mentioned that, geometrically, the curves are homothetic branches of rectangular hyperbolas in the (π_{ab}, β_c) -plane of the parameter domain centered at $(1 - \pi_\emptyset^0, 1)$.

If identifiability is not conceived as a property of a probabilistic model, but rather as a property of its parameters, then the above results indicate that in the considered BLIM the parameters β_c and π_{ab} cannot be identified, while this is not the case for the remaining parameters π_\emptyset , β_a , and β_b .

Forward- and backward-gradedness of the underlying knowledge structure, however, is not the only source of non-identifiability of a BLIM. Consider the knowledge structure $\mathcal{K} = \{\emptyset, \{a, b\}, \{a, c\}, Q\}$ on $Q = \{a, b, c\}$ (taken from Example 4.39), which has neither of these properties. However there exists a linear dependence relation among the columns of the Jacobian matrix. The critical property in the present knowledge structure \mathcal{K} is the fact that $\mathcal{K}_b^* = \{Q \setminus \{c\}\}$ and $\mathcal{K}_c^* = \{Q \setminus \{b\}\}$. In general if, in the otherwise arbitrary knowledge structure \mathcal{K} , for any $p, q \in Q$ we have $\mathcal{K}_p^* = \{Q \setminus \{q\}\}$ and $\mathcal{K}_q^* = \{Q \setminus \{p\}\}$, then

$$\begin{aligned} & (1 - \beta_p - \eta_p) \cdot \pi_{Q \setminus \{p\}} \cdot \frac{\partial f_R}{\partial \beta_p}(\theta) - \pi_{Q \setminus \{p\}} \cdot (1 - \sum_{L \in \mathcal{K}_p^*} \pi_L) \cdot \frac{\partial f_R}{\partial \pi_{Q \setminus \{p\}}}(\theta) \\ &= (1 - \beta_q - \eta_q) \cdot \pi_{Q \setminus \{q\}} \cdot \frac{\partial f_R}{\partial \beta_q}(\theta) - \pi_{Q \setminus \{q\}} \cdot (1 - \sum_{L \in \mathcal{K}_q^*} \pi_L) \cdot \frac{\partial f_R}{\partial \pi_{Q \setminus \{q\}}}(\theta) \end{aligned} \quad (4.93)$$

holds for all $R \in \mathcal{R}^*$ (Heller, 2017, proposition 2).

The cases considered in this section (backward- and forward-gradedness, as well as Equation (4.93)) need not cover all instances of non-identifiability that may crop up. There are, however, conjectures on the parameter trade-offs that may occur with any kind of non-identifiability (Heller, 2017). From the examples considered so far, one might surmise that the π -parameters need to be involved in any parameter trade-off. However, Heller (2017, proposition 5) showed that if the knowledge structure is both forward- and backward-graded in an item q , and the parameter constraints $\pi_K = \pi_{K \cup \{q\}}$ hold for all $K \in \mathcal{K}_{\bar{q}}$, then

$$\frac{\partial f_R}{\partial \beta_q}(\theta) = -\frac{\partial f_R}{\partial \eta_q}(\theta) \quad (4.94)$$

for all $R \in \mathcal{R}^*$ with $q \in R$. Because the assumed constraints result in a subset of the parameter domain with null measure, this suggests the following weaker formulation.

Conjecture 4.65 *The trade-offs between parameters due to non-identifiability of a BLIM involve some of the π -parameters together with at least one of either or both of the β - or η -parameters almost everywhere in the parameter space.*

Another conjecture refers to “equally informative items”: items p and q are *equally informative* if they co-occur in all knowledge states of \mathcal{K} , i.e., $\mathcal{K}_p = \mathcal{K}_q$

holds. It was conjectured by Spoto *et al.* (2013) that the parameters $\beta_p, \beta_q, \eta_p, \eta_q$ attached to equally informative items p and q are never part of the parameter trade-offs due to non-identifiability. Using a similar argument as above, Heller (2017, proposition 6) showed that this need not be the case, again on a subset of the parameter domain with null measure. Thus, the conjecture has to be revised as follows.

Conjecture 4.66 *Equally informative items are not involved in trade-offs due to non-identifiability of a BLIM almost everywhere in the parameter space.*

The relevance of Conjecture 4.66 lies in the fact that non-identifiable BLIMs are likely to be encountered in practical applications. The knowledge structures that have received most attention suffer from non-identifiability due to forward- or backward-gradedness, among them the learning spaces. Establishing Conjecture 4.66 would thus open a possible way for restoring identifiability: it would then suffice to extend the knowledge domain in such a way that each item comes with at least one other equally informative item. Formal proofs of both Conjectures 4.65 and 4.66, however, are still missing.

Non-identifiability is not only a property of a BLIM that one intends to avoid for theoretical reasons; it may cause serious problems to knowledge assessment, which tries to single out a person's knowledge state. Using a Bayesian rule, assessment is commonly based on the posterior distribution on the knowledge states given the observed response pattern, i.e., by considering

$$\mathbb{P}(K | R) = \frac{\mathbb{P}(R | K) \cdot \mathbb{P}(K)}{\mathbb{P}(R)} \quad (4.95)$$

with knowledge state K and response pattern R . It could be decided that the assessed state is identified with the maximum *a posteriori* (MAP) estimate, i.e., the state that maximizes $\mathbb{P}(K | R)$. Whenever the parameters of the BLIM are not determined uniquely then this holds true for the MAP estimate, too.

Example 4.67 Consider the already discussed fair BLIM defined on the knowledge structure $\mathcal{K} = \{\emptyset, \{a, b\}, Q\}$ on $Q = \{a, b, c\}$ (see Example 4.64). Figure 4.7 illustrates the trade-off between components π_{ab} and β_c of the parameter vector $\theta = (\pi_\emptyset, \pi_{ab}, \beta_a, \beta_b, \beta_c)$. Consider the parameter values $\theta_1 = (0.4, 0.12, 0.1, 0.1, 0.05)$ and $\theta_2 = (0.4, 0.03, 0.1, 0.1, 0.2)$, which predict the same distribution on the response patterns, i.e., $f(\theta_1) = f(\theta_2)$. This means that we cannot distinguish between θ_1 and θ_2 on empirical grounds. However, the posterior distribution of the states given the response pattern will be dependent upon the particular parameter value, which might affect the assessed knowledge state. Considering the MAP estimate for the response pattern $R = \{a, b\}$, for example, reveals that the state $K = \{a, b\}$ results given that $\text{BLIM}(\mathcal{K})$ is based on θ_1 , while the state $K = \{a, b, c\}$ results when referring to θ_2 .

Software implementing numerical methods for determining the rank of the Jacobian matrix of the prediction function of a BLIM (of moderate size) is available

(Heller & Wickelmaier, 2013; Stefanutti *et al.*, 2012). In view of Proposition 4.16, they allow for recognizing non-identifiable BLIMs, and the parameter trade-offs they entail.

4.10.1 Trade-off Dimensions

The BLIM essentially consists of three different types of parameters: lucky guesses η , careless errors β , and knowledge state probabilities π . Non-identifiability is a consequence of trade-off among these three types of parameters. Stefanutti *et al.* (2012) analyze identifiability of the BLIM by distinguishing trade-offs occurring among parameters of the same type from those that occur across different parameter types. In this way they distinguish three orders of trade-offs:

- (i) *first-order trade-offs*: parameters involved in this type of trade-off are all of the same type (e.g., a trade-off among lucky guess probabilities of different items);
- (ii) *second-order trade-offs*: these are trade-offs involving parameters of two different types (e.g., a trade-off between the careless error probability of some item and the probability of some knowledge state);
- (iii) *third-order trade-offs*: involve parameters of all three types (e.g., a trade-off involving the lucky guess of an item, the careless error of another item, and the probability of some knowledge state).

These three classes of trade-offs can be studied separately through a suitable decomposition of the Jacobian matrix into the three different submatrices P , B , and E , whose entries are given by Equations (4.73), (4.74), and (4.75), respectively. Thus, P is the $(2^{|Q|} - 1) \times (|\mathcal{K}| - 1)$ matrix of the first derivatives of the prediction function f with respect to the knowledge state probabilities π ; B is the $(2^{|Q|} - 1) \times |Q|$ matrix of the first derivatives of f w.r.t. the careless error parameters β ; E is the $(2^{|Q|} - 1) \times |Q|$ matrix of the first derivatives of f w.r.t. the lucky guess parameters η . The whole Jacobian matrix is reconstructed by

$$\mathbf{J}_f(\theta) = [P|B|E]. \quad (4.96)$$

First-order trade-offs among parameters of the same type occur when one or more submatrices P , B , E are not full-rank. For instance, $\text{rk}(B) < |Q|$ would indicate the presence of trade-offs among the β parameters of some of the items. In this case, the total number of first-order trade-off dimensions for the careless error parameters is measured by $\text{null}(B) = |Q| - \text{rk}(B)$. Similar considerations apply to the two submatrices P and E . Therefore, the total number of first-order trade-off dimensions for all parameter types is given by

$$\text{null}(P) + \text{null}(B) + \text{null}(E). \quad (4.97)$$

It is still unclear whether first-order trade-offs can occur in a BLIM. It is clear that they cannot occur when the analysis is restricted to the matrix P (see the end of this section), but, concerning each of the two matrices B and E , there are no conclusive results.

Second-order trade-offs are detected by considering pairs of submatrices. For $X, Y \in \{P, B, E\}$ there are in total three possible pairs of submatrices, each of which gives rise to a compound submatrix of the form $[X|Y]$. The number of second-order trade-off dimensions contained in submatrix $[X|Y]$ is obtained by subtracting the number of first-order trade-off dimensions from the total number of trade-off dimensions in $[X|Y]$. Indicating with $td_2([X|Y])$ the number of second-order trade-off dimensions contained in $[X|Y]$, we obtain

$$td_2([X|Y]) = \text{null}([X|Y]) - \text{null}(X) - \text{null}(Y), \quad (4.98)$$

which can also be expressed in terms of ranks. If X and Y have c and d columns, respectively, we get

$$\begin{aligned} td_2([X|Y]) &= c + d - \text{rk}([X|Y]) - c + \text{rk}(X) - d + \text{rk}(Y) \\ &= \text{rk}(X) + \text{rk}(Y) - \text{rk}([X|Y]). \end{aligned} \quad (4.99)$$

An example of a second-order trade-off is provided by the curves in the vector field of Figure 4.7. Each of the curves is the geometrical representation of a mapping $t \mapsto \theta(t)$. The mapping involves parameters of two types (the β_q 's and the π_K 's in the example). As one of the two parameters is increased, the other one can be decreased in a way that leaves unchanged the probability distribution $\theta(t)$ on the response patterns.

Second-order trade-off dimensions arise specifically in forward-graded and backward-graded knowledge structures. In a backward-graded knowledge structure one has one trade-off dimension for every single item in which the structure is backward-graded. The trade-off is of a second order because it only involves β_q and π_K parameters. An analogous situation arises with the forward-graded knowledge structures, with the only difference that the parameters involved in the trade-off are the π_K and η_q parameters.

Finally, the number $td_3([P|B|E])$ of third-order trade-off dimensions satisfies

$$\begin{aligned} td_3([P|B|E]) &= -\text{rk}(P) - \text{rk}(B) - \text{rk}(E) \\ &\quad + \text{rk}([P|B]) + \text{rk}([P|E]) + \text{rk}([B|E]) - \text{rk}([P|B|E]). \end{aligned} \quad (4.100)$$

Example 4.68 Consider the structure

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b, c\}, \{a, b, d\}, \{a, b, c, d\}\}. \quad (4.101)$$

It is neither forward- nor backward-graded. Computing the rank of the Jacobian matrix for $\text{BLIM}(\mathcal{K})$ at randomly sampled points θ in the parameter space \mathcal{D} gave $\text{rk}(\mathbf{J}_f(\theta)) = 10$. This means that, with a total number of $2^Q - 2|Q| - |\mathcal{K}| - 1 = 13$ parameters, the Jacobian is not full-rank and hence $\text{BLIM}(\mathcal{K})$ is not locally identifiable at the sampled points. In particular, the null space of the Jacobian is almost surely three-dimensional. Moreover, the number of trade-off dimensions is $td_3 = 3$ for this BLIM. The decomposition of the Jacobian matrix \mathbf{J}_f into the three submatrices P, B, E leads to the results summarized in Table 4.2, which consists of three different columns and seven rows. The first column (Submatrices) is a list of all

Table 4.2 *Decomposition of the Jacobian matrix from Example 4.68 into the three submatrices B , E , and P and their concatenations, and submatrix rank analysis.*

Submatrices	No. param.	Rank
B	4	4
E	4	4
P	5	5
$[B E]$	8	8
$[P B]$	9	8
$[P E]$	9	8
$[P B E]$	13	10

the possible ways of concatenating the three submatrices B , E , and P into larger submatrices (seven in total). The second column of the table (No. param.) displays the number of columns (BLIM's parameters) in each of the submatrices listed in the first column. For instance, the submatrix $[B|E]$ has four β_q parameters plus five π_K parameters, for a total of eight parameters). The third column (Rank) indicates the rank of each of the listed submatrices.

Three of the seven submatrices listed in Table 4.2 are not full-rank, namely $[P|B]$, $[P|E]$, and $[P|B|E]$. By applying Equations (4.99) and (4.100) it can be concluded that: (i) there are no first-order trade-off dimensions (all three matrices B , E , and P are full-rank); (ii) there are two second-order trade-off dimensions, one for $[P|B]$ and one for $[P|E]$; (iii) there is exactly one third-order trade-off dimension, involving all three types of parameters for some items and some states. Stefanutti *et al.* (2012) describe a numerical method for establishing what specific parameters are involved in what trade-off dimensions.

For a finite set Q , the family of all possible knowledge structures is the collection of all subsets of the power set 2^Q containing at least the empty set and Q . Then this collection can be theoretically partitioned into the set of all knowledge structures for which the BLIM is identifiable, and the set of all those knowledge structures for which it is not. Considering knowledge structures belonging to this last collection, they might differ from one another concerning the types of trade-offs occurring among the model parameters. There is a lot to study about similarities and differences in this respect. A basic result is provided in Stefanutti *et al.* (2012), which holds for all knowledge structures on a finite set Q . It essentially claims that there are no first-order trade-offs among the probabilities of the knowledge states.

Proposition 4.69 *Let \mathcal{K} be any knowledge structure on the finite set Q and, for $\theta \in \mathcal{D}$, let $J_f(\theta) = [P|B|E]$ be the Jacobian matrix of the prediction function f of $BLIM(\mathcal{K})$. Then P is full-rank.*

The question of whether a similar proposition holds, for each of the two submatrices B and E , remains open.

4.11 Latent Class Models

Following Schrepp (2005) and Ünlü (2011) we now introduce latent class models (Goodman, 1974; Lazarsfeld, 1955; Lazarsfeld & Henry, 1968) in a way that emphasizes their relationship with the BLIM. In particular, we restrict ourselves to the dichotomous case, where each observed variable can take only two values. Suppose we observe the joint occurrence of b variables Y_1, Y_2, \dots, Y_b , with Y_i taking value in the set $\{0, 1\}$. We collect the frequency of any of the 2^b patterns described by $Y_1 = v_1, Y_2 = v_2, \dots, Y_b = v_b$ where $v_i \in \{0, 1\}$ for $i = 1, 2, \dots, b$. With $I = \{1, 2, \dots, b\}$, the pattern is also characterized as the subset $R = R(v) = \{i \in I : v_i = 1\}$ of I . We write \mathcal{R} for the set of all the 2^b patterns. In our application we observe students taking a test with b items, setting $Y_i = 1$ if the answer to item i is correct.

Latent class analysis aims at explaining the observed frequencies of the patterns by the following mechanism. There is a population of reference (in our application, the population of students), partitioned into c latent classes K_1, K_2, \dots, K_c , with π_j the (hidden) probability that an individual belongs to class K_j ($j = 1, 2, \dots, c$). There exist also (hidden) numbers α_{ij} , with $0 \leq \alpha_{ij} \leq 1$, where $i = 1, 2, \dots, b$ and $j = 1, 2, \dots, c$, with the interpretation that α_{ij} is the conditional probability that the variable Y_i receives value 1 from an individual in class j . The basic assumptions of the latent class model (LCM) are: (i) all the individuals belonging to the same class K_j share the same probability for the occurrence of any given response pattern; (ii) moreover, this probability results from an independence assumption on the variables Y_i for the individuals in class j . In formal terms, LCM sets the probability of occurrence of the pattern $Y_1 = v_1, Y_2 = v_2, \dots, Y_b = v_b$ equal to

$$f_v(\pi, \alpha) = \sum_{j=1}^c \pi_j \prod_{i \in \{1, 2, \dots, b\}: v_i=0} (1 - \alpha_{ij}) \prod_{i \in \{1, 2, \dots, b\}: v_i=1} \alpha_{ij}, \quad (4.102)$$

where (π_1, \dots, π_c) is the vector of class probabilities, $\alpha = (\alpha_{ij})$ is the $b \times c$ matrix of conditional probabilities α_{ij} , and $v = (v_1, v_2, \dots, v_b)$ is the description of the response pattern. In other words, the predicted probability of the pattern R , a subset of $I = \{1, 2, \dots, b\}$, is

$$f_R(\pi, \alpha) = \sum_{j=1}^c \pi_j \prod_{i \in I \setminus R} (1 - \alpha_{ij}) \prod_{i \in R} \alpha_{ij}. \quad (4.103)$$

Thus, $f = (f_R)_{R \in \mathcal{R}}$ denotes the prediction function of the LCM. From $0 \leq \pi_i$ and $0 \leq \alpha_{ij} \leq 1$, there directly follows $0 \leq f_R(\pi, \alpha)$. Moreover, it is an exercise to show that there holds $\sum_{R \in \mathcal{R}} f_R(\pi, \alpha) = 1$ for any π and α with $\sum_{j=1}^c \pi_j = 1$.

Here is a more specific description of our LCM as a probabilistic model. With redundancy, the parameters of the LCM model consist of b probabilities π_j for the

latent classes, and $b \times c$ conditional probabilities α_{ij} of the observable values v_i , given the latent classes K_j . Since the latent class probabilities must sum to 1, there are only $c - 1$ non-redundant latent class parameters π_j in the model; we decide to leave out π_c . Thus, in its non-redundant form, the LCM has a total of $c - 1 + b \cdot c$ parameters. For a LCM in which c distinct latent classes are hypothesized, the assumption is made that every class K_j has a positive probability, otherwise the number of hypothesized latent classes would simply be less than c (this assumption holds as far as the postulate of existence of a certain latent class coincides with the assumption that it has a positive probability in the population). Thus we assume

$$\pi_j > 0, \quad j = 1, 2, \dots, c - 1. \quad (\text{LC1})$$

This implies that every π_j lies in the open interval $]0, 1[$. Moreover, the fact that π is a probability distribution over \mathcal{R} and $\pi_c > 0$ implies

$$\sum_{j=1}^{c-1} \pi_j < 1. \quad (\text{LC2})$$

If there are no reasons for excluding certain outcomes of a variable Y_i , given a latent class K_j , then it may be retained that every outcome value for Y_i is observable, given class K_j , meaning that

$$0 < \alpha_{ij} < 1, \quad i = 1, 2, \dots, b, \quad j = 1, 2, \dots, c. \quad (\text{LC3})$$

On the other hand, suppose that theoretical reasons or empirical evidence suggest $\alpha_{ij} = 0$ for some i . Then there would be no reason to keep parameter α_{ij} in the model. That is, setting α_{ij} to zero is essentially the same as removing that parameter from the model, with the consequence of changing the parameter space itself.

Given all of this, the parameter space \mathcal{D} of the LCM consists of the vectors

$$\theta = (\pi_1, \pi_2, \dots, \pi_{c-1}, \alpha_{1,1}, \alpha_{2,1}, \dots, \alpha_{b,1}, \alpha_{1,2}, \dots, \alpha_{b,c}) \quad (4.104)$$

in $]0, 1[^{c-1+b \cdot c}$ that satisfy Conditions (LC1), (LC2), and (LC3); we denote such a vector also by the shorthand (π, α) . Then \mathcal{D} is both open and convex in $\mathbb{R}^{c-1+b \cdot c}$.

Now for the outcome space. The LCMs described in this section are aimed at predicting the outcome of an experiment, regarded as the set of the (joint) realizations of the dichotomous random variables Y_i in the experiment itself (where $i = 1, 2, \dots, b$). There are thus 2^b theoretically observable results of the experiment. As we saw just after Equation (4.103), the prediction function f assigns, for any $\theta = (\pi, \alpha)$ given in the parameter domain, a probability distribution on the collection \mathcal{C} of all response patterns. The outcome space would thus be the collection of all the probability distributions on the set \mathcal{R} . We again meet redundancy: we may forget the probability value of one pattern, say the pattern I ; we set again $\mathcal{R}^* = \mathcal{R} \setminus \{Q\}$. Then the outcome space \mathcal{O} consists of all points $(y_R)_{R \in \mathcal{R}^*}$ satisfying

$$0 \leq y_R \leq 1, \quad R \in \mathcal{R}^*, \quad (4.105)$$

$$\sum_{R \in \mathcal{R}^*} y_R \leq 1. \quad (4.106)$$

Any BLIM happens¹² to be a “submodel” of some LCM, more precisely: for any knowledge structure \mathcal{K} , it holds that the $\text{BLIM}(\mathcal{K}) = (\mathcal{D}_B, f_B, \mathcal{O}_B)$ is a “submodel” of some LCM $(\mathcal{D}_L, f_L, \mathcal{O}_L)$, where “submodel” requires the existence of some injective mapping $g : \mathcal{D}_B \rightarrow \mathcal{D}_L$ and some bijective mapping $h : \mathcal{O}_B \rightarrow \mathcal{O}_L$ such that f_L restricted to $g(\mathcal{D}_B)$ equals $h \circ f_B \circ g^{-1}$. To prove this, assume the domain Q (that is $\cup \mathcal{K}$) equals $\{1, 2, \dots, b\}$. Defining the latent classes to be exactly the knowledge states and the items to be the same for both models, we form a LCM $(\mathcal{D}_L, f_L, \mathcal{O}_L)$. Next we define a mapping $g : \mathcal{D}_B \rightarrow \mathcal{D}_L$ by letting the image of any parameter point $\theta = (\pi, \beta, \eta)$ in \mathcal{D}_B be the point $g(\theta) = (\pi, \alpha)$ in \mathcal{D}_L with, first, the same parameters π_j as those of the BLIM and, second,

$$\alpha_{ij} = \begin{cases} 1 - \beta_i & \text{if } i \in K_j, \\ \eta_i & \text{if } i \notin K_j. \end{cases} \quad (4.107)$$

The outcome spaces of the BLIM and the LCM are identical, so we let $h : \mathcal{O}_B \rightarrow \mathcal{O}_L$ be the identity mapping. To check the condition on the prediction mappings, we compute the image by f_L of a point $g(\pi, \beta, \eta)$. In view of Equation (4.103), $f_L(g(\pi, \beta, \eta))$ assigns to the response pattern R the probability

$$\sum_{K \in \mathcal{K}} \pi_K \left(\prod_{q \in K \setminus R} \beta_q \right) \left(\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right) \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \left(\prod_{q \in R \setminus K} \eta_q \right).$$

This happens to be just a rewriting of the core equation of $\text{BLIM}(\mathcal{K})$. So, we have established that $\text{BLIM}(\mathcal{K})$ is a submodel of the LCM just constructed.

We must however emphasize that knowledge structure theory (KST) is not subsumed to the theory of LCMs. Indeed, KST takes into account an additional organization of the set of classes, captured by the knowledge structure itself (this explains why we can, for instance, take advantage of an assumption like backward-gradedness).

To summarize, the BLIM is a latent class model where the probability of $Y_i = v_i$ (the dichotomous response to item i) depends on the latent variable (knowledge state) K_j in a simple way. As shown in Equation (4.107), the probability is totally specified by the validity (or invalidity) of the formula $i \in K_j$. For any fixed random variable Y_i , the set \mathcal{K} of latent classes is thus partitioned into the two collections \mathcal{K}_i and $\mathcal{K}_{\bar{i}}$ (containing the classes K for which $i \in K$, resp., $i \notin K$).

The BLIM has smaller format than the corresponding dichotomous LCM, in that α_{ij} only depends on i . Its main advantage thus lies in the reduced number of parameters. However, this particular restriction prevents a direct generalization to general LCMs of the results obtained on (local) identifiability for the BLIM. There exist probabilistic knowledge structure models, more general than the BLIM, that specify a less trivial dependence of the observable responses on the patterns from the latent knowledge states (de Chiusole *et al.*, 2013). However, results on local identifiability of such models are still missing.

¹² As sketched in footnote 4 of Ünlü (2006).

Table 4.3 *For a given value c of the number of latent classes, the minimum value of the number b of observed variables for which Proposition 4.20 implies that the LCM is quantitatively testable.*

c	Minimum b	c	Minimum b
1	3	10	7
2	4
3	4	15	7
4	5	16	8
5	5
6	6	28	8
7	6	29	9
8	6
9	6	51	9

We turn back to general LCMs. A latent class model $(\mathcal{D}, f, \mathcal{O})$ is testable if its prediction range $f(\mathcal{D})$ is a proper subset of the outcome space of the model. Because of Conditions (LC1) and (LC3), every response pattern appears for each parameter point (π, α) with a positive probability $f_R(\pi, \alpha)$ and this is sufficient for assuring that $f(\mathcal{D})$ is a proper subset of \mathcal{O} , so that our LCMs are testable.

Concerning quantitative testability of the LCM $(\mathcal{D}, f, \mathcal{O})$, tools from Section 4.3.4 are directly applicable. Notice that the prediction function f is polynomial, thus also C^∞ and analytic. According to Proposition 4.20, quantitative testability holds when the parameter domain is of dimension strictly less than the outcome space. Here the latter inequality amounts to $c - 1 + b \cdot c < 2^b - 1$. Thus when $c < 2^b/(b + 1)$ the LCM is quantitatively testable. The upper bound on c only depends on the total number b of observable variables Y_i . For instance, with only one observable variable, the LCM is not quantitatively testable. With two or three variables, at most one latent class is admitted in a quantitatively testable LCM. With four observable variables the maximum number of classes is three, and so on. This means that, without any further restrictions, the LCMs with at least two latent classes that are quantitatively testable comprise at least four dichotomous variables. Now let us take the opposite point of view. For small values of c , Table 4.3 gives a value of b for which we are sure that the LCM is quantitatively testable – the LCM is also quantitatively testable for all larger values of b . The LCM could be quantitatively testable also for lower values of b : by Propositions 4.19 and 4.20, this happens exactly if the rank $\text{rk}(J_f)$ of the Jacobian matrix of the prediction function f is always less than $2^b - 1$.

We now show that the dichotomous LCM $(\mathcal{D}, f, \mathcal{O})$ is never identifiable. The reason is very simple: take any point (π, α) in \mathcal{D} satisfying for two distinct classes j, j' the following condition for all i in I :

$$\alpha_{ij} = \alpha_{ij'}. \quad (4.108)$$

Then in Equation (4.102) π_j and $\pi_{j'}$ are multiplied by equal expressions. Hence we may regroup the terms involving π_j and $\pi_{j'}$ as one term involving $\pi_j + \pi_{j'}$. So we conclude that adding a small amount to π_j and subtracting the same amount from $\pi_{j'}$ will not change the value of the prediction function. The argument shows that local identifiability fails at any parameter point $(\boldsymbol{\pi}, \boldsymbol{\alpha})$ of the LCM $(\mathcal{D}, f, \mathcal{O})$. It is interesting that the above argument does not apply to the BLIM seen as a submodel of a LCM as earlier. Indeed, the resulting parameters $\boldsymbol{\alpha}$ never satisfy Equation (4.108), which is here equivalent to $1 - \beta_i = \eta_i$ for all items i in $K_j \Delta K_{j'}$. This directly contradicts Condition (C5), which reads $b_i + \eta_i < 1$.

There results an interesting question: for a given LCM $(\mathcal{D}, f, \mathcal{O})$, characterize the predicted points which are identifiable. In particular, decide whether there are “many” of them as compared to the predicted points (where “many” should be given a precise meaning). Lazarsfeld and Henry (1968) expose nice results in this direction: for instance, their theorems 1 and 3 (on pp. 56 and 64) provide sufficient conditions for the identifiability of a predicted point.

As a final remark, we stress that our point of view in the whole chapter was focused on the theoretical aspects of probabilistic models. Statistical assessment of models is another topic; for LCM, see for instance chapter 4 in Lazarsfeld and Henry (1968).

4.12 Conclusions

This chapter investigated fundamental properties of probabilistic models, which are generally referred to as testability and identifiability. According to definitions in Bamber and van Santen (1985, 2000), both notions come in a weak and a strong version. While a model is testable whenever not all outcomes that may be observed are predicted by the model, it is quantitatively testable if predicted outcomes are sparse, such that the probability of sampling them by chance is zero. Global identifiability refers to a situation where any distinct parameter points predict distinct observed outcomes, so that the parameter values are uniquely determined by the observation. In the weaker version of local identifiability this one-to-one relation need not hold everywhere in the parameter domain, but in a neighborhood of a parameter point only.

The practical significance of testability is quite obvious, as it characterizes the falsifiability of a theory. If many points of an outcome space are actually lying in the prediction range of a model, however, we may not be impressed by observing a predicted outcome. This evaluation changes if a quantitatively testable model is considered. Its highly precise predictions make it very unlikely that the model fits the data by chance.

For a probabilistic model, the lack of (local) identifiability has severe implications for its application. As shown above, non-identifiability leads to trade-offs between parameters. Although this does not limit the global fit to data which may be achieved, it will impair parameter estimation. The likelihood of given data is

identical for all parameter points mapped onto the same outcome by the prediction function. In maximum likelihood estimation this will inflate the standard errors of the estimates for the parameters involved in trade-offs, and iterative procedures may not even converge.

It should be clear that the estimates obtained for these parameters cannot be interpreted, because they only represent a single configuration of values from the (possibly infinitely) many constituting the trade-off, all of which are empirically indistinguishable. Notice that non-identifiability as characterized here is a property of the probabilistic model *per se*, and independent of any data. However, even if a model is identifiable, the information in the data may not suffice to distinguish between certain parameter points. In the literature this is sometimes referred to as a lack of “practical identifiability,” or of “empirical identifiability.” Insufficient information in the data is a problem that comes on top of the identifiability issues considered in this chapter.

For the basic local independence model (BLIM) we have seen that non-identifiability occurs as a consequence of particular structural properties of the underlying knowledge structure. Whenever a BLIM is not identifiable, the above-mentioned problems in parameter estimation will most likely be encountered. Then the estimates of (some of) the parameter values cannot be interpreted, and thus it may not be possible to assess a unique knowledge state given an individual’s response pattern. In one of the conjectures presented above, it was surmised that including equally informative items (co-occurring in all the knowledge states) into the knowledge domain may be an option to essentially restore identifiability. This refers to still another concept of identifiability which, to the best of our knowledge, has not been considered so far, and which requires a precise definition and formal investigations. It might be called “essential identifiability.” Assuming its outcome space has positive measure, a probabilistic model is “essentially identifiable” if the set of parameter points at which it is not identifiable is of zero measure. Besides the mentioned conjecture on BLIMs, many positive results of point identifiability (as for LCMs) should imply essential identifiability. It may thus be promising to investigate this concept further.

4.13 Bibliographical Notes

The notions of testability and identifiability of probabilistic models appear in various disciplines, such as econometrics (Fisher, 1966), statistics (Lehmann & Casella, 1998), and psychology (Bamber & van Santen, 1985, 2000).

The first publication on knowledge structure theory is Doignon and Falmagne (1985). For more details, we refer the reader to Doignon and Falmagne (2016), a chapter in the first volume of the Handbook.

The two particular probabilistic models we investigate here figure either implicitly (CRM) or explicitly (BLIM) in Doignon and Falmagne (1999). The BLIM appeared for the first time in Falmagne and Doignon (1988), the CRM received

its name in Doignon (2017). The results about the CRM are taken from Doignon (2017), those about the BLIM from several papers. Spoto *et al.* (2012) introduced the structural properties of forward- and backward-gradedness and formulated first results on the induced non-identifiability of the BLIM, which were extended by Spoto *et al.* (2013). Routines for numerically determining the rank of the Jacobian matrix of the BLIM were provided by Stefanutti *et al.* (2012) for MATLAB (2011), and by Heller and Wickelmaier (2013) for the R Software Environment for Statistical Computing (R Core Team, 2016). Heller (2017) presented further theoretical results on local non-identifiability, showing in particular that there are instances not due to forward- or backward-gradedness.

Latent class analysis was formalized by Lazarsfeld (see for instance Lazarsfeld, 1955). A good exposition is Lazarsfeld and Henry (1968), with identifiability defined and studied in chapter 3. A paper addressing identifiability is Goodman (1974). For more recent papers and books on latent class models, see, for instance, the website www.john-uebersax.com/stat/faq.htm

Acknowledgments

The authors thank Ali Ünlü for providing them with information on latent class models, and Laurent Fourny for suggesting many improvements to a preliminary version of their text.

References

- Bamber, D., & van Santen, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, 29, 443–473.
- Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, 44, 20–40.
- Birkhoff, G. (1937). Rings of sets. *Duke Mathematical Journal*, 3, 443–454.
- Bolt, D. (2007). The present and future of IRT-based cognitive diagnostic models (ICDMs) and related methods. *Journal of Educational Measurement*, 44, 377–383.
- Boothby, W. M. (1975). *An introduction to differentiable manifolds and Riemannian geometry*. Pure and Applied Mathematics, No. 63. New York, NY: Academic Press.
- de Chiusole, D., Stefanutti, L., Anselmi, P., & Robusto, E. (2013). Assessing parameter invariance in the BLIM: Bipartition models. *Psychometrika*, 78, 710–724.
- DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, 35, 8–26.
- DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, 44, 285–291.
- Dieudonné, J. (1960a). *Foundations of modern analysis*. Pure and Applied Mathematics, Vol. X, vol. I. New York, NY: Academic Press.

- Dieudonné, J. (1960b). *Foundations of modern analysis*. Pure and Applied Mathematics, Vol. X, vol. III. New York, NY: Academic Press.
- Doignon, J.-P. (2017). *A Correct Response Model in Knowledge Space Theory*. Manuscript in revision.
- Doignon, J.-P., & Falmagne, J.-Cl. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23, 175–196.
- Doignon, J.-P., & Falmagne, J.-Cl. (1999). *Knowledge spaces*. Berlin: Springer-Verlag.
- Doignon, J.-P., & Falmagne, J.-Cl. (2016). Knowledge spaces and learning spaces. In W. H. Batchelder, H. Colonius, E. N. Dzhafarov, & J. Myung (eds.), *New handbook of mathematical psychology* (pp. 274–321). Cambridge: Cambridge University Press.
- Falmagne, J.-C., & Doignon, J.-P. (1988). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, 41, 1–23.
- Fisher, F. M. (1966). *The identification problem in econometrics*. New York, NY: McGraw-Hill.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215–231.
- Grünbaum, B. (2003). *Convex polytopes*, second edition. Graduate Texts in Mathematics, Vol. 221. Prepared and with a preface by V. Kaibel, V. Klee and G. M. Ziegler. New York, NY: Springer.
- Haertel, E. H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement*, 8, 333–346.
- Haertel, E. H. (1989). Using restricted latent class models to map skill structure of achievement items. *Journal of Educational Measurement*, 26, 301–321.
- Heller, J. (2017). Identifiability in probabilistic knowledge structures. *Journal of Mathematical Psychology*, 77, 46–57.
- Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, 42, 49–56.
- Heller, J., Ünlü, A., & Albert, D. (2013). Skills, competencies and knowledge structures. In J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, & X. Hu (eds.), *Knowledge spaces: Applications in education* (pp. 229–242). New York, NY: Springer-Verlag.
- Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 80, 995–1019.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258–272.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement*, Vol. 1. New York, NY: Academic Press.
- Lang, S. (1993). *Real and functional analysis*, third edition. Graduate Texts in Mathematics, Vol. 142. Berlin: Springer-Verlag.
- Lang, S. (1997). *Undergraduate analysis*, second edition. Undergraduate Texts in Mathematics. Berlin: Springer-Verlag.
- Lazarsfeld, P. F. (1955). Recent developments in latent structure analysis. *Sociometry*, 18, 391–403.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. New York, NY: Houghton Mifflin.

- Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*, second edition. Springer Texts in Statistics. New York, NY: Springer-Verlag.
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: John Wiley.
- MATLAB. (2011). *MATLAB version 7.13.0 (R2011b)*. The Mathworks, Inc., Natick, Massachusetts.
- R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Roberts, F. S. (1979). *Measurement theory, with applications to decision-making, utility, and the social sciences*. Reading, MA: Addison-Wesley.
- Rudin, W. (1976). *Principles of mathematical analysis*, third edition. New York, NY: McGraw-Hill.
- Sard, A. (1942). The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society*, 48, 883–890.
- Sard, A. (1958). Images of critical sets. *Annals of Mathematics: Second Series*, 68, 247–259.
- Schrepp, M. (2005). About the connection between knowledge structures and latent class models. *Methodology*, 1, 93–103.
- Shapiro, A., & Browne, M. W. (1983). On the investigation of local identifiability: A counterexample. *Psychometrika*, 48, 303–304.
- Smith, P. L. (1998). Attention and luminance detection: A quantitative analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 105–133.
- Spoto, A., Stefanutti, L., & Vidotto, G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *Journal of Mathematical Psychology*, 56, 248–255.
- Spoto, A., Stefanutti, L., & Vidotto, G. (2013). Considerations about the identification of forward- and backward-graded knowledge structures. *Journal of Mathematical Psychology*, 57, 249–254.
- Stanley, R. P. (1986). Two poset polytopes. *Discrete and Computational Geometry*, 1, 9–23.
- Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing the local identifiability of probabilistic knowledge structures. *Behavior Research Methods*, 44, 1197–1211.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 667–680.
- Suppes, P., & Zinnes, J. L. (1963). *Basic measurement theory*, third edition. Vol. 1, pp. 1–76. New York, NY: Wiley.
- Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51, 337–350.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287–305.
- Ünlü, U. (2006). Estimation of careless error and lucky guess probabilities for dichotomous test items: A psychometric application of a biometric latent class model with random effects. *Journal of Mathematical Psychology*, 50, 309–328.
- Ünlü, U. (2011). A note on the connection between knowledge structures and latent class models. *Methodology*, 7, 63–67.
- Wald, A. (1950). *Statistical inference in dynamic economic models*. Cowles Commission Monographs, Vol. 10. New York, NY: John Wiley & Sons.
- Ziegler, G. M. (1998). *Lectures on polytopes*, revised edition. Berlin: Springer-Verlag.

5 Quantum Models of Cognition and Decision

Jerome R. Busemeyer and Peter D. Kvam

5.1	What is Quantum Cognition?	186
5.2	Mathematical Background	186
5.2.1	Hilbert Space	187
5.2.2	Linear Operators and Subspaces	189
5.2.3	Basis Vectors and Unitary Operators	191
5.2.4	Eigenvectors and Eigenvalues of Matrices	194
5.2.5	Tensor Product Spaces	195
5.3	Quantum versus Classical Probability Theories	196
5.3.1	Events	196
5.3.2	System State	197
5.3.3	State Revision	198
5.3.4	Commutative Events	198
5.3.5	Non-commutative Events	199
5.3.6	Observables and the Uncertainty Relation	201
5.3.7	Generalized Measurements and Density Operators	202
5.3.8	Psychological Applications	204
5.3.9	Conjunction Fallacy	204
5.3.10	Question Order Effects	205
5.3.11	Disjunction Effect	207
5.4	Markov versus Quantum Dynamics	208
5.4.1	System State	208
5.4.2	Transitions between States	209
5.4.3	Measurements	211
5.4.4	Generalized Measurements and Density Operators	212
5.4.5	Psychological Applications	213
5.5	Concluding Comments	217
5.5.1	Open Questions and Problems	218
5.5.2	Extensions and Alternatives to Quantum Probability	218
	References	219

5.1 What is Quantum Cognition?

Quantum cognition concerns the application of the mathematical principles from quantum theory for the purpose of developing probabilistic models of human behavior.¹ Of course, quantum theory was originally developed for physical applications. However, after Dirac (1930) and von Neumann (1932) axiomatized the theory, the abstract principles could be lifted out of the physical context, and then applied to other fields such as psychology or economics. This is analogous to many of the formalisms that arose in physics but then spread out to other fields. For example, much of the recurrent neural network theory is based on dynamical principles that were originally formulated for classical physics.

Why apply quantum theory to cognition? We try to answer this question more rigorously in the remainder of this chapter, but for now we simply point to its success in a variety of applications including judgment (Aerts & Aerts, 1994; Busemeyer, Pothos, Franco, & Trueblood, 2011; Franco, 2009; Pothos, Busemeyer, & Trueblood, 2013; Wang & Busemeyer, 2013), decision-making (Bordley & Kadane, 1999; Busemeyer, Wang, & Townsend, 2006; Khrennikov & Haven, 2009; Lambert-Mogiliansky, Zamir, & Zwirn, 2009; La Mura, 2009; Pothos & Busemeyer, 2009; Trueblood & Busemeyer, 2011; Yukalov & Sornette, 2010), conceptual combinations (Aerts, 2009; Aerts & Gabora, 2005; Blutner, 2008), memory (Brainerd, Wang, & Reyna, 2013; Bruza *et al.*, 2009), and perception (Atmanspacher, Filk, & Romer, 2004; Conte *et al.*, 2009). Several review articles (Pothos & Busemeyer, 2013; Wang *et al.*, 2013) and books (Busemeyer & Bruza, 2012; Ivancevic & Ivancevic, 2010; Khrennikov, 2010) provide a summary of this new program of research.

This chapter is organized into three main sections. The first reviews some of the mathematical background required to understand quantum probability theory in a rigorous manner. The second describes the structural and measurement part of quantum theory in comparison with classical probability theory. The third part describes the dynamical part of quantum theory in comparison with classical Markov theories.

5.2 Mathematical Background

Both cognitive and neuroscientists employ multidimensional vector spaces to represent conceptual features, memory traces, or activation patterns across neural networks. The branch of mathematics concerned with vector spaces is linear algebra (for finite dimensional spaces) and functional analysis (for infinite dimensional spaces). Quantum theory is essentially a probability theory built on a foundation of vector spaces. For this reason, quantum theory provides a natural mathematical framework for representing probabilities for events that occur in

¹ This differs from quantum brain theory (Hammeroff, 1998), because quantum cognition does not rely on any particular theory regarding neural substrates.

cognitive or neural models. Below we review some of the mathematical concepts and definitions required to rigorously understand quantum probability theory.

In the following, we use the Dirac notation for vectors and inner products and linear operators. This is important for two reasons. We do not immediately use coordinates and matrices to represent vectors or linear operators, because the latter require first choosing a basis, and we wish to define the basic concepts in a general manner that holds for any basis. This allows us to move more freely from one basis to another, which we often need to do in quantum theory. Second, the Dirac notation facilitates identifying relationships between the abstract mathematical concepts.

5.2.1 Hilbert Space

As we describe in Section 5.3, quantum theory is based on the mathematical concept of a Hilbert space, which is a vector space defined on a complex field endowed with an inner product that is complete with respect to the metric induced by the inner product. There are a lot of mathematical concepts in that one sentence. Let us unpack it.

First, a Hilbert space is a vector space. A vector space is a set \mathcal{H} of abstract elements (called vectors). In cognitive neuroscience, for example, we can represent an activation pattern across a large number of neural units as a vector. Using the Dirac notation, vectors in \mathcal{H} are symbolized by “kets” such as $|X\rangle$, $|Y\rangle$, and $|Z\rangle$. A vector space includes elements produced by the operation of vector addition $|X\rangle + |Y\rangle$ which is commutative $|X\rangle + |Y\rangle = |Y\rangle + |X\rangle$ and associative $|Z\rangle + (|X\rangle + |Y\rangle) = (|Z\rangle + |X\rangle) + |Y\rangle$. A vector space contains a special element, denoted 0 such that $|X\rangle + 0 = |X\rangle$ and for every vector $|X\rangle$ there is a corresponding vector $|-X\rangle$ such that $|X\rangle + (-|X\rangle) = 0$. A vector space also includes elements produced by multiplication by a scalar α , which is an element in the field of complex numbers, to form a new vector $\alpha \cdot |X\rangle$. Scalar multiplication is associative $(\alpha \cdot \beta) \cdot |X\rangle = \alpha \cdot (\beta \cdot |X\rangle)$ and distributive $(\alpha + \beta) |X\rangle = \alpha \cdot |X\rangle + \beta \cdot |X\rangle$, $\alpha \cdot (|X\rangle + |Y\rangle) = \alpha \cdot |X\rangle + \beta \cdot |Y\rangle$, and there is a scalar, denoted 1 such that $1 \cdot |X\rangle = |X\rangle$.

The dimension of a vector space is determined by the number of linearly independent vectors that exist in the space. A vector $|Y\rangle$ is linearly *dependent* on a set of vectors $|X_i\rangle$, $i = 1, 2, \dots, N$ if there exists a non-zero N -tuple of coefficients such that $|Y\rangle = \sum_{i=1, N} \alpha_i \cdot |X_i\rangle$, and otherwise it is linearly *independent*. If the space has N linearly independent vectors, but not $N+1$ linearly independent vectors, then the space is finite with dimension equal to N . Otherwise it is infinite dimensional. For example, the human brain is estimated to have approximately 86 billion neurons. If we wished to represent the activation pattern across all of these neurons, then we would need at most this many dimensions, but this large number still remains finite.

Second, the vector space is defined on a complex field. Complex fields are commonly used in cognitive neuroscience for signal processing, such as with the Fourier transformation. The field of complex numbers are all pairs of real numbers $\alpha = (a, b)$ in the complex plain, where a is called the real part and b is called

the imaginary part. Often a complex number is written as $\alpha = (a, b) = a + b \cdot i$ with $i^2 = -1$. The imaginary number i has two functions: it keeps track of the real and imaginary parts by applying i to the imaginary part, and it is convenient for defining addition and multiplication of complex numbers. The addition of complex numbers is defined by $\alpha + \beta = (a + b \cdot i) + (c + d \cdot i) = (a + c) + (b + d) \cdot i$, and the multiplication of complex numbers is defined by $\alpha \cdot \beta = (a + b \cdot i)(c + d \cdot i) = (a \cdot c - b \cdot d) + (a \cdot d + b \cdot c) \cdot i$. The conjugate of a complex number is defined as $\alpha^* = a - b \cdot i$ and the squared magnitude of a complex number equals $|\alpha|^2 = \alpha \cdot \alpha^* = a^2 + b^2$, and its absolute value then equals $|\alpha| = \sqrt{a^2 + b^2}$. We can determine a phase θ such that $\cos(\theta) = \frac{a}{|\alpha|}$ and $\sin(\theta) = \frac{b}{|\alpha|}$ and then we can redefine the complex number as $\alpha = a + b \cdot i = |\alpha| \cdot (\cos(\theta) + i \cdot \sin(\theta))$. It is often more convenient to work with the complex exponential definition $e^{i\theta} = \cos(\theta) + i \cdot \sin(\theta)$, so that $\alpha = a + b \cdot i = |\alpha| \cdot e^{i\theta}$. This way we can use the rules for complex exponentials to perform calculations such as $\alpha \cdot \beta = |\alpha| \cdot e^{i\theta_1} \cdot |\beta| \cdot e^{i\theta_2} = |\alpha \cdot \beta| \cdot e^{i(\theta_1+\theta_2)}$.

Third, the Hilbert space is endowed with an inner product. For example, in cognitive neuroscience, the correlated activation of two different networks of neurons can be measured by their inner product. An inner product is a mapping of a pair of vectors ($|X\rangle, |Y\rangle$) into the scalar field of complex numbers. Using Dirac notation, the scalar produced by the inner product for the pair ($|X\rangle, |Y\rangle$) is denoted as $\langle X|Y\rangle$. An inner product is defined by the following properties:

1. the inner product for the pair (X, X) equals $\langle X|X\rangle \geq 0$ and $\langle X|X\rangle = 0 \iff |X\rangle = 0$;
2. the inner product for the pair (Y, X) equals $\langle Y|X\rangle = \langle X|Y\rangle^*$;
3. the inner product for the pair ($|Z\rangle, \alpha \cdot |X\rangle + \beta \cdot |Y\rangle$) equals $\alpha \cdot \langle Z|X\rangle + \beta \cdot \langle Z|Y\rangle$.

Together these imply the following anti-linear property for inner products: the inner product for the pair ($(\alpha \cdot |X\rangle + \beta \cdot |Y\rangle, |Z\rangle)$) equals $\alpha^* \cdot \langle X|Z\rangle + \beta^* \cdot \langle Y|Z\rangle$. As discussed below, inner products are used to measure lengths of vectors and similarities between vectors. In particular, if $\langle X|Y\rangle = 0$, then the vectors are orthogonal.

Inner products can be viewed as linear functionals. In general, a functional is a map f from a vector in \mathcal{H} into a scalar. A linear functional I is a functional that satisfies the property of linearity:

$$I(\alpha |X\rangle + \beta |Y\rangle) = \alpha \cdot I(|X\rangle) + \beta \cdot I(|Y\rangle).$$

For example, the definite integral of a continuous function is a linear functional. The inner product for a pair ($|W\rangle, |X\rangle$), where we keep the first element $|W\rangle$ in the pair fixed, and we allow the second element $|X\rangle$ to vary arbitrarily in \mathcal{H} , defines a functional $\langle W|$ which satisfies the property of linearity $\langle W|(\alpha |X\rangle + \beta |Y\rangle) = \alpha \cdot \langle W|X\rangle + \beta \cdot \langle W|Y\rangle$. The linear functional $\langle W|$ is called a “bra,” and for this reason, the inner product $\langle W|X\rangle$ is called a “bra-ket.” Each vector in \mathcal{H} can be mapped into a linear functional in this manner. The set of all linear functionals forms another vector space called the dual space of \mathcal{H} and so $\langle W|$ is a vector in this dual space. The operation that maps a vector $|W\rangle$ from \mathcal{H} into a linear functional $\langle W|$ in the dual space is called the adjoint operation, which is denoted as $|W\rangle^\dagger = \langle W|$. The

inverse operation maps a linear functional in the dual space back into a vector in the Hilbert space, which is denoted $\langle W \rangle^\dagger = |W\rangle$. Using the adjoint operation, we can define the inner product as $|X\rangle^\dagger |Y\rangle = \langle X|Y\rangle$. The adjoint operation is anti-linear because $(\alpha \cdot |W\rangle)^\dagger = \alpha^* \langle W|$ and $(\alpha^* \cdot \langle W|)^\dagger = \alpha \cdot |W\rangle$, as required to satisfy the properties of inner products.

The inner product induces a norm on the Hilbert space, denoted as $\| |X\rangle \|$, which maps a vector into a length (a finite non-negative real number). A norm has the following properties:

1. $\| |X\rangle \| \geq 0$ and $\| |X\rangle \| = 0 \longleftrightarrow |X\rangle = 0$;
2. $\| \alpha \cdot |X\rangle \| = |\alpha| \cdot \| |X\rangle \|$, and
3. $\| |X\rangle + |Y\rangle \| \leq \| |X\rangle \| + \| |Y\rangle \|$.

We wish to define a norm that is based on the algebraic properties of inner products. Therefore, the norm induced by the inner product is defined as $\| |X\rangle \| = \sqrt{\langle X|X\rangle}$. A normalized vector is one that has unit length, $\| |X\rangle \| = 1$.

A norm also allows us to define a metric on a space. A metric space is simply a set of points that has a distance function d that measures the distance between pairs of points. If $|X\rangle$, $|Y\rangle$, and $|Z\rangle$ are arbitrary points in the metric space (not necessarily vectors from a vector space), then the distance function satisfies the following properties:

1. $d(|X\rangle, |Y\rangle) \geq 0$ and $d(|X\rangle, |Y\rangle) = 0 \longleftrightarrow |X\rangle = |Y\rangle$,
2. symmetry $d(|X\rangle, |Y\rangle) = d(|Y\rangle, |X\rangle)$, and
3. the triangle inequality $d(|X\rangle, |Z\rangle) \leq d(|X\rangle, |Y\rangle) + d(|Y\rangle, |Z\rangle)$.

If the space has a norm, then the distance function is induced by the norm as follows: $d(|X\rangle, |Y\rangle) = \| |X\rangle - |Y\rangle \|$. When working with inner products, the inner product induces a distance function defined as $d(|X\rangle, |Y\rangle) = \| |X\rangle - |Y\rangle \| = \sqrt{(|X\rangle - |Y\rangle)^\dagger (|X\rangle - |Y\rangle)}$.

Finally, a Hilbert space is a complete inner product space. In order to define completeness, we first need to define a Cauchy sequence. A sequence of points $|X_i\rangle$, $i = 1, 2, \dots$, from a metric space is a Cauchy sequence if for every $\epsilon > 0$ there is an N (depending on ϵ) such that $d(|X_i\rangle, |X_j\rangle) < \epsilon$ for every $(i, j) > N$. A metric space is complete if every Cauchy sequence converges, that is, has a limit in the metric space. In sum, a metric space can be complete or not complete. A complete metric space endowed with a norm is called a Banach space. A Banach space endowed with an inner product is called a Hilbert space. Quantum cognition researchers generally work with finite dimensional vector spaces defined on a complex field endowed with an inner product and these spaces are always complete, so they are Hilbert spaces.

5.2.2 Linear Operators and Subspaces

Cognitive scientists and neuroscientists often employ networks that transform an input pattern of activation (i.e., an input vector) into an output pattern of activation (i.e., an output vector). In many cases these transformations are linear (e.g., linear

transformation of features used in support vector machines). Mathematically, a linear operator \mathbf{L} is a transformation that takes a vector $|X\rangle$ in the Hilbert space \mathcal{H}_1 and maps it into another vector $|Y\rangle$ in another Hilbert space \mathcal{H}_2 , $\mathbf{L}|X\rangle = |Y\rangle$, with the following property:

$$\mathbf{L}(\alpha|X\rangle + \beta|Y\rangle) = \alpha \cdot \mathbf{L}|X\rangle + \beta \cdot \mathbf{L}|Y\rangle.$$

An example of a linear operator is the derivative of a function.

If the mapping is one to one, then the inverse operator exists: $\mathbf{L}^{-1}|Y\rangle = \mathbf{L}^{-1}\mathbf{L}(|X\rangle) = |X\rangle$. Assuming that the domain of a second linear operator \mathbf{M} matches the range of the first linear operator \mathbf{L} , then the two linear operators can be applied successively to produce a sequence of transformations, for example, $\mathbf{ML}|X\rangle = \mathbf{M}|Y\rangle = |Z\rangle$. If the order does not matter, so that the commutator $[\mathbf{M}, \mathbf{L}] = \mathbf{ML} - \mathbf{LM} = 0$, then the operators *commute*. For example, if the inverse \mathbf{L}^{-1} exists, then it commutes with \mathbf{L} so that $\mathbf{LL}^{-1} = \mathbf{L}^{-1}\mathbf{L}$. However, very often linear operators do not commute, so that in many cases the commutator $[\mathbf{M}, \mathbf{L}] = \mathbf{ML} - \mathbf{LM} \neq 0$. Non-commutativity is a key concept in quantum theory.

Linear operators can be bounded or unbounded. For example, the derivative is not necessarily bounded. A linear operator is continuous if and only if it is bounded. However, for finite dimensional spaces, linear operators are always bounded and therefore always continuous.

Recall that each Hilbert space has a dual space whose vectors are linear functionals. We can also apply linear operators, say for example \mathbf{L}^\dagger , to these vectors in the dual space to produce the linear transformation $\langle W|\mathbf{L}^\dagger = \langle V|$, which corresponds to the transformation $\langle V|^\dagger = (\langle W|\mathbf{L}^\dagger)^\dagger = \mathbf{L}|W\rangle = |V\rangle$ in the Hilbert space \mathcal{H} . The operator \mathbf{L}^\dagger is called the adjoint of \mathbf{L} if for all $(|X\rangle, |W\rangle)$ in \mathcal{H} ,

$$(L|W\rangle)^\dagger|X\rangle = |W\rangle^\dagger(L^\dagger|X\rangle).$$

An important type of linear operator is the Hermitian operator that satisfies the property $\mathbf{L}^\dagger = \mathbf{L}$. A positive operator is a Hermitian operator that also satisfies $\langle X|\mathbf{L}|X\rangle > 0$ for all $|X\rangle$.

There are some other important types of linear operators. One is the identity operator, denoted \mathbf{I} , with the property that $\mathbf{I}|X\rangle = |X\rangle$. Another is the zero operator, $\mathbf{0}|X\rangle = 0$. Another special kind of linear operator is called an outer product. Using Dirac notation, the outer product of the pair of vectors $(|X\rangle, |Y\rangle)$ from \mathcal{H} is denoted as $|X\rangle\langle Y|$ and this operator produces the following transformation: $(|X\rangle\langle Y|)|X\rangle = |X\rangle\langle Y|W\rangle = \langle Y|V\rangle \cdot |X\rangle$.

In cognitive science, a concept, such as “cat,” is often represented as the presence of some subset of features in a multidimensional space. Mathematically, this can be represented as a subspace of the Hilbert space. Each subspace, such as \mathcal{H}_A , is a vector space formed by a subset of the Hilbert space, $\mathcal{H}_A \subset \mathcal{H}$. Corresponding to each subspace \mathcal{H}_A , there is a linear operator, \mathbf{P}_A , called the orthogonal projector for the subspace. To produce an orthogonal projection, the projector \mathbf{P}_A must be

1. idempotent, $\mathbf{P}_A \mathbf{P}_A = \mathbf{P}_A^2 = \mathbf{P}_A$, and
2. Hermitian $\mathbf{P}_A^\dagger = \mathbf{P}_A$.

The projector maps vectors in the Hilbert space \mathcal{H} onto the subspace \mathcal{H}_A . The resulting projection $\mathbf{P}_A |X\rangle = |X_A\rangle$ has the following properties: if we define $|Z\rangle = |X\rangle - |X_A\rangle$, then

$$|X\rangle = \mathbf{P}_A |X\rangle + (I - \mathbf{P}_A) |X\rangle = |X_A\rangle + Z$$

and

$$\langle Z | X_A \rangle = \langle X | (I - \mathbf{P}_A^\dagger) \mathbf{P}_A | X \rangle = \langle X | (\mathbf{P}_A - \mathbf{P}_A) | X \rangle = 0.$$

Thus, the projector decomposes each $|X\rangle$ into its projection and the orthogonal residual. The simplest kind of subspace is a ray, that is, the set of points produced by an arbitrary scalar multiplication of a vector, $\alpha \cdot |X\rangle$ with $\| |X\rangle \| = 1$. The projector for this subspace is defined by the outer product $\mathbf{P}_X = |X\rangle \langle X|$ which produces the projection $\mathbf{P}_X |Y\rangle = |X\rangle \langle X| Y = \langle X | Y \rangle \cdot |X\rangle$. A *partition* or orthogonal decomposition of the N -dimensional Hilbert space is defined as a set of projectors $\{\mathbf{P}_i, i = 1, M \leq N\}$ such that $\mathbf{P}_i \mathbf{P}_j = \mathbf{0}$ for all $i \neq j$ and $\sum_{i=1,M} \mathbf{P}_i = \mathbf{I}$.

5.2.3 Basis Vectors and Unitary Operators

So far we have not used any matrices to represent vectors and linear operators. To introduce matrices, we need to choose a particular basis for representing vectors in the Hilbert space. Psychologically, choosing a basis can be interpreted as selecting a particular point of view or perspective for defining features that are used to answer questions about events. Assume that the dimension of the Hilbert space is some finite number N (e.g., 86 billion for all the neurons in the human brain). This implies that there are at most N linearly independent vectors that span the Hilbert space. Given any N linearly independent vectors that span the space, we can use the Gram–Schmidt algorithm to produce a set of N orthogonalized and normalized (orthonormal) vectors that span the space. This set of N orthonormal vectors $\mathcal{V} = \{|V_1\rangle, \dots, |V_N\rangle\}$ forms an orthonormal basis for representing the space. The outer product $P_i = |V_i\rangle \langle V_i|$ is the projector that projects vectors in \mathcal{H} onto the ray spanned by the basis vector $|V_i\rangle$. The collection of N projectors $\{\mathbf{P}_{Vi}, i = 1, N\}$ provides an orthogonal decomposition of the Hilbert space \mathcal{H} : $P_{Vi} P_{Vj} = \mathbf{0}$ for $i \neq j$ and $\sum \mathbf{P}_{Vi} = \mathbf{I}$. Any vector $|X\rangle$ can be expressed in terms of the \mathcal{V} basis as follows

$$|X\rangle = \mathbf{I} |X\rangle = \sum \mathbf{P}_{Vi} |X\rangle = \sum |V_i\rangle \langle V_i| X \rangle = \sum \psi_i |V_i\rangle,$$

with $\psi_i = \langle V_i | X \rangle$. Then the $N \times 1$ matrix $\boldsymbol{\psi} = [\psi_i]$ provides the coordinates of the vector $|X\rangle$ with respect to the \mathcal{V} basis. In particular, the basis vector $|V_j\rangle$ is represented by the $N \times 1$ matrix $V_j = [\delta(i,j)]$, for rows $i = 1, \dots, N$, where $\delta(i,i) = 1$ and $\delta(i,j) = 0$ for row $i \neq j$. Note that if $\boldsymbol{\psi} = [\psi_i]$ is the $N \times 1$ matrix representation of the vector $|X\rangle$ in the \mathcal{V} basis, then the $1 \times N$ matrix representation of the adjoint $\langle X|$ is the conjugate transpose $\boldsymbol{\psi}^\dagger = [\psi_i^*]$. Also note that if $\boldsymbol{\psi} = [\psi_i]$

is a $N \times 1$ column matrix, then we use the $1 \times N$ row matrix $\psi^T = [\psi_i]$ to represent the transpose of the matrix (without conjugation).

In the N -dimensional space, the linear operator \mathbf{L} mapping from \mathcal{H} to \mathcal{H} can be represented by the $N \times N$ matrix of inner products $L = [L_{ij}]$, $L_{ij} = \langle V_i | \mathbf{L} | V_j \rangle$. Alternatively, the linear operator can be expressed in terms of the matrix coordinates of basis \mathcal{V} as $\mathbf{L} = \sum L_{ij} \cdot |V_i\rangle \langle V_j|$.

Suppose $|Y\rangle = \mathbf{L}|X\rangle = \sum \phi_i |V_i\rangle$ in the \mathcal{V} basis, so that $\phi = [\phi_i]$ is the $N \times 1$ matrix representing $|Y\rangle$ in the \mathcal{V} basis. Then the corresponding matrix transformation is given by

$$\begin{aligned}\phi_i &= \langle V_i | \mathbf{L} | X \rangle \\ &= \langle V_i | \sum_j L_{ij} \cdot |V_i\rangle \langle V_j| \sum_k \psi_k |V_k\rangle, \\ &= \sum_j L_{ij} \cdot \psi_j \rightarrow \\ \phi &= L \cdot \psi.\end{aligned}$$

Applying this result twice, we get the matrix representation for a sequence of two linear transformations: if $|Z\rangle = \mathbf{M}|Y\rangle = \mathbf{ML}|X\rangle = \sum \gamma_i |V_i\rangle$ and M is the matrix representation of \mathbf{M} with respect to the \mathcal{V} basis, then $\gamma = M \cdot \phi = M \cdot L \cdot \psi$.

Using the coordinates for the \mathcal{V} basis, we can compute the inner product between the pair $(|X\rangle, |Y\rangle)$ by the matrix formula

$$\begin{aligned}\langle X | Y \rangle &= \left(\sum \psi_i^* \langle V_i | \right) \left(\sum \phi_j | V_j \rangle \right) \\ &= \sum \sum \psi_i^* \phi_j \langle V_i | V_j \rangle \\ &= \sum \psi_i^* \phi_i = \psi^\dagger \cdot \phi,\end{aligned}$$

where $\psi = [\psi_i]$ is a $N \times 1$ column matrix, $\psi^\dagger = [\psi_i^*]$ is a $1 \times N$ row matrix, and $\phi = [\phi_i]$ is a $N \times 1$ column matrix.

With respect to the \mathcal{V} basis, the projector $\mathbf{P}_{Vj} = |V_j\rangle \langle V_j|$ that projects on the ray spanned by $|V_j\rangle$ is represented by a diagonal matrix M_{Vj} with $\delta(i,j)$ as a diagonal element in row i , so that $M_{Vj} \cdot \psi = [0 \ \dots \ \psi_j \ \dots \ 0]^T$ (where the superscript T represents the simple transpose of a matrix). The projector for the three-dimensional subspace $\mathbf{P}_{Vi} + \mathbf{P}_{Vj} + \mathbf{P}_{Vk}$ spanned by the set $\{|V_i\rangle, |V_j\rangle, |V_k\rangle\}$ is represented in the \mathcal{V} basis by $M_{Vi} + M_{Vj} + M_{Vk} = \text{diag}[0, \dots, 1, 0 \dots 0, 1, 0 \dots 1, 0 \dots]$ with zeros along the diagonal except for the value 1 in rows (i, j, k) .

In quantum theory, we frequently need to change the basis. Psychologically, this corresponds to changing the point of view or perspective for defining the features used to evaluate questions. The \mathcal{V} basis is one of infinitely many bases that we could choose to represent the Hilbert space \mathcal{H} . In general, a unitary operator \mathbf{U} is a linear operator that is defined by the following properties: $\mathbf{U}^\dagger \mathbf{U} = \mathbf{U} \mathbf{U}^\dagger = \mathbf{I}$. Unitary transformations are defined as transformations that preserve lengths and inner products: suppose $|V\rangle = \mathbf{U}|X\rangle$ and $|W\rangle = \mathbf{U}|Y\rangle$; then $\langle V | W \rangle = \langle X | \mathbf{U}^\dagger \mathbf{U} | Y \rangle =$

$\langle X|Y \rangle$. We can transform from the \mathcal{V} basis to a new orthonormal basis $\mathcal{W} = \{\mathbf{U}|V_1\rangle = |W_1\rangle, \dots, \mathbf{U}|V_N\rangle = |W_N\rangle\}$ using the unitary matrix $\mathbf{U} = \sum |W_i\rangle\langle V_i|$. Note that it is not necessary to change *all* basis vectors by this transformation, because we could set $|W_k\rangle = |V_k\rangle$ for some of the indices. Using the \mathcal{W} basis to represent vectors, we have the new representation $|X\rangle = \sum \omega_i |W_i\rangle$. In this case, the $N \times 1$ column matrix $\boldsymbol{\omega} = [\omega_i]$, where $\omega_i = \langle W_i|X\rangle$ provides the coordinates of the vector $|X\rangle$ with respect to the \mathcal{W} basis. In the N -dimensional space, the linear operator \mathbf{L} mapping from \mathcal{H} to \mathcal{H} can be represented by the $N \times N$ matrix $M = [M_{ij}]$, $M_{ij} = \langle W_i|\mathbf{L}|W_j\rangle$. Thus, changing the basis changes the coordinate system used to represent vectors and matrices.

We can represent the unitary operator \mathbf{U} in terms of the basis \mathcal{V} by the $N \times N$ unitary matrix $U = [u_{ij}]$, with $u_{ij} = \langle V_i|\mathbf{U}|V_j\rangle$. Likewise, the adjoint of the unitary operator corresponds to the Hermitian transpose of the unitary matrix U^\dagger . Then we can express the transformation of the coordinates of $|X\rangle$ for the \mathcal{V} basis into the coordinates of $|X\rangle$ for the \mathcal{W} basis as follows:

$$\begin{aligned}\omega_i &= \langle W_i|X\rangle = \langle W_i|\mathbf{I}|X\rangle \\ &= \left\langle V_i|U^\dagger \sum |V_j\rangle\langle V_j|X\right\rangle \\ &= \sum \left\langle V_i|U^\dagger|V_j\right\rangle \cdot \psi_j \rightarrow \\ \boldsymbol{\omega} &= U^\dagger \cdot \boldsymbol{\psi} \\ \boldsymbol{\psi} &= U \cdot \boldsymbol{\omega}.\end{aligned}$$

The last two equations determine how to change the coordinates of the same vector $|X\rangle$ from one basis \mathcal{V} to another \mathcal{W} . We frequently need to use these formulas for changing coordinates in quantum theory. Note that the inner product does not depend on the coordinate system that is used to represent vectors, because if $|X\rangle$ is represented by the $N \times 1$ column matrices $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$ with respect to the \mathcal{V} and \mathcal{W} bases, and $|Y\rangle$ is represented by the $N \times 1$ column matrices $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$ with respect to the \mathcal{V} and \mathcal{W} bases, then $\boldsymbol{\gamma} = U^\dagger \cdot \boldsymbol{\psi}$ and $\boldsymbol{\eta} = U^\dagger \cdot \boldsymbol{\phi}$ and $\langle X|Y \rangle = \boldsymbol{\gamma}^\dagger \cdot \boldsymbol{\eta} = \boldsymbol{\psi}^\dagger \cdot U \cdot U^\dagger \cdot \boldsymbol{\phi} = \boldsymbol{\psi}^\dagger \cdot \boldsymbol{\phi}$. Also note that if $L = [\langle V_i|L|V_j\rangle]$ is the matrix representation of the linear operator \mathbf{L} with respect to the basis \mathcal{V} , and $M = [\langle W_i|\mathbf{L}|W_j\rangle]$ is the matrix representation of the same linear operator \mathbf{L} with respect to the basis \mathcal{W} , then $M_{ij} = [\langle W_i|\mathbf{L}|W_j\rangle] = [\langle V_i|U^\dagger LU|V_j\rangle]$, which implies $M = U^\dagger \cdot L \cdot U$.

The projectors for different subspaces are also related by a unitary operator. Suppose that subspace \mathcal{H}_A is spanned by a subset \mathcal{V}_A of basis vectors in \mathcal{V} so that the projector is defined as $\mathbf{P}_A = \sum_{j \in A} |V_j\rangle\langle V_j|$, and suppose that subspace \mathcal{H}_B is spanned by a subset \mathcal{W}_B of vectors in \mathcal{W} so that the projector is defined as $\mathbf{P}_B = \sum_{j \in B} |W_j\rangle\langle W_j|$. Also suppose the basis for \mathcal{W} is obtained by a unitary transformation \mathbf{U} of the basis for \mathcal{V} . Then we have the relation $\mathbf{P}_B = \sum_{j \in B} |W_j\rangle\langle W_j| = \mathbf{U} \left(\sum_{j \in B} |V_j\rangle\langle V_j| \right) \mathbf{U}^\dagger$. If we choose to express all of these operators as matrices in terms of the \mathcal{V} basis, then U is the unitary matrix that changes from one coordinate system to the other; \mathbf{P}_A corresponds to a diagonal matrix M_A with ones

located at indices corresponding to the basis vectors in \mathcal{V}_A and zeros otherwise; and \mathbf{P}_B corresponds to the matrix $U \cdot M_B \cdot U^\dagger$ where M_B corresponds to a diagonal matrix with ones located at indices corresponding to the basis vectors in \mathcal{V}_B and zeros otherwise. Note that if $\mathbf{U} = \mathbf{I}$, so that the same basis \mathcal{V} is used to span both subspaces, then \mathbf{P}_B commutes with \mathbf{P}_A , and otherwise they do not.

5.2.4 Eigenvectors and Eigenvalues of Matrices

Eigen decompositions are frequently used in cognitive science for the purpose of dimension reduction and data compression. They serve many different fundamental roles in quantum theory. An eigenvector of a linear operator \mathbf{L} is defined as a (non-zero) vector $|V\rangle$ that satisfies the eigen equation $\mathbf{L}|V\rangle = \lambda \cdot |V\rangle$ for some scalar λ called the eigenvalue. If the Hilbert space is finite, there is always at least one non-zero eigenvalue and at most N unique eigenvalues. Given that we are working in a finite dimensional space, and we have chosen a basis for representing the space, we can do all of our work with matrices and transform from one coordinate system to another using unitary matrices. Then we can compute eigenvectors and eigenvalues using the matrix representations.

Hermitian operators play a special role in quantum theory. Suppose \mathbf{H} is a Hermitian linear operator so that $\mathbf{H}^\dagger = \mathbf{H}$. Then we can decompose this operator into N orthonormal eigenvectors as follows:

$$\mathbf{H} = \sum_{i=1,N} \lambda_i \cdot |V_i\rangle \langle V_i|,$$

where $|V_i\rangle$ is an eigenvector with real eigenvalue λ_i . If all of the eigenvalues are unique, then the decomposition is unique. If there are repeated eigenvalues, then the decomposition can still be formed, but it is no longer unique.

In particular, consider a projector $\mathbf{P}_A = \sum_{j=1,M} |V_j\rangle \langle V_j|$ that projects vectors onto subspace $\mathcal{H}_A \subset \mathcal{H}$ spanned by basis vectors $\mathcal{V}_A = \{|V_i\rangle, i = 1, M\}$, which is a subset of the orthonormal basis $\mathcal{V} = \{|V_i\rangle, i = 1, N\}$ with $N > M$. This linear operator has only two eigenvalues, (0, 1). Each basis vector in \mathcal{V}_A is an eigenvector of \mathbf{P}_A with an eigenvalue equal to 1; and each of the remaining basis vectors in \mathcal{V} , orthogonal to \mathcal{V}_A , is an eigenvector with an eigenvalue equal to 0. However, \mathcal{V}_A is only one of many possible bases for spanning the space \mathcal{H}_A . Suppose that \mathbf{U} is a unitary operator that commutes with \mathbf{P}_A , and $\mathcal{W}_A = \{|W_i\rangle = \mathbf{U}|V_i\rangle, i = 1, M\}$, is another basis for spanning \mathcal{H}_A . Then

$$\begin{aligned} \sum_{j=1,M} |W_j\rangle \langle W_j| &= \mathbf{U} \sum_{j=1,M} |V_j\rangle \langle V_j| \mathbf{U}^\dagger \\ &= \mathbf{U} \mathbf{P}_A \mathbf{U}^\dagger = \mathbf{P}_A \mathbf{U} \mathbf{U}^\dagger = \mathbf{P}_A. \end{aligned}$$

A positive operator has eigenvalues that are all positive real valued. The matrix of a positive operator is positive definite with positive eigenvalues. A positive definite matrix $H = V \cdot \Lambda \cdot V^\dagger$ has a square root defined as

$$\sqrt{H} = V \cdot \sqrt{\Lambda} \cdot V^\dagger$$

where $\sqrt{\Lambda} = \text{diag} [\cdots \sqrt{\lambda_i} \cdots]$ because $\sqrt{H} \cdot \sqrt{H} = V \cdot \sqrt{\Lambda} \cdot V^\dagger V \cdot \sqrt{\Lambda} \cdot V^\dagger = H$. Using a positive definite matrix, we can define a more general expression for an inner product. For example, if α and β are arbitrary $N \times 1$ column matrices taken from C^N , and H is a positive definite matrix, we can define an inner product between α and β by the matrix product $\langle \alpha | \beta \rangle = \alpha^\dagger \cdot H \cdot \beta$. This satisfies all the properties of an inner product because it can be rewritten as $\alpha^\dagger \cdot H \cdot \beta = (\alpha^\dagger \cdot V \cdot \sqrt{\Lambda}) (\sqrt{\Lambda} \cdot V^\dagger \beta) = (\sqrt{\Lambda} \cdot V^\dagger \alpha)^\dagger (\sqrt{\Lambda} \cdot V^\dagger \beta)$.

More generally, functions of operators can be defined using the eigen decomposition of an operator. In particular, any unitary operator can be expressed as an exponential function of a Hermitian operator:

$$U = \exp(-i \cdot H) = \sum_{i=1,N} e^{(-i \cdot \lambda_i)} |V_i\rangle \langle V_i|.$$

The latter provides an important formula for constructing the unitary matrices that we need for changing coordinates of systems.

Another important function of an operator is the trace of an operator. If L is a linear operator mapping from \mathcal{H} to \mathcal{H} , and L is the square matrix representation of the operator with respect to some basis $\mathcal{V} = \{|V_i\rangle, i = 1, N\}$, then its trace is defined as

$$\text{Tr}[L] = \sum_{i=1,N} \langle V_i | L | V_i \rangle = \sum_{i=1,N} L_{ii},$$

that is, the sum of the diagonal elements. The trace is linear because $\text{Tr}[\alpha \cdot L + \beta \cdot M] = \sum_{i=1,N} \langle V_i | \alpha \cdot L + \beta \cdot M | V_i \rangle = \alpha \cdot \sum_{i=1,N} \langle V_i | L | V_i \rangle + \beta \cdot \sum_{i=1,N} \langle V_i | M | V_i \rangle = \alpha \cdot \text{Tr}[L] + \beta \cdot \text{Tr}[M]$. The trace is cyclic, so that if L, M are two square matrices, then $\text{Tr}[L \cdot M] = \sum \sum (L_{ij} \cdot M_{ji}) = \text{Tr}[M \cdot L]$. The latter property guarantees that the trace of an operator does not depend on the basis used to define the matrix, because if L is a matrix expressed in coordinates of $\mathcal{V} = \{|V_i\rangle, i = 1, N\}$ and M is a matrix expressed in coordinates of $\mathcal{W} = |W_i\rangle = U |V_i\rangle$, so that $M = U^\dagger \cdot L \cdot U$, then $\text{Tr}[L] = \text{Tr}[M] = \text{Tr}[U^\dagger (L \cdot U)] = \text{Tr}[L \cdot U \cdot U^\dagger] = \text{Tr}[L]$. It is handy to note that the trace of an outer product equals an inner product: $\text{Tr}[|X\rangle \langle Y|] = \langle Y | X \rangle$.

5.2.5 Tensor Product Spaces

Tensor products have frequently appeared in models of cognitive science, especially in the area of memory modeling for binding multiple representations into a single memory trace. A tensor product space is a new Hilbert space formed by combining other Hilbert spaces. Suppose \mathcal{H}_A is a Hilbert space spanned by a set of basis vectors $\mathcal{V}_A = \{|V_i\rangle, i = 1, M\}$, and suppose \mathcal{H}_B is a separate Hilbert space spanned by a set of basis vectors $\mathcal{V}_B = \{|W_i\rangle, i = 1, N\}$. Then the tensor product space is an $N \cdot M$ -dimensional space, denoted $\mathcal{H}_A \otimes \mathcal{H}_B$, spanned by basis vectors $\mathcal{V}_{AB} = \{|V_i\rangle \otimes |W_j\rangle, i = 1, M; j = 1, N\}$. If $|X\rangle = \sum \psi_i \cdot |V_i\rangle$ is a vector in \mathcal{H}_A , and $|Y\rangle = \sum \phi_i \cdot |W_i\rangle$ is a vector in \mathcal{H}_B , then

$$|X\rangle \otimes |Y\rangle = \sum \sum (\psi_i \cdot \phi_j) (|V_i\rangle \otimes |V_j\rangle)$$

is a vector in the tensor product space. Furthermore, if ψ is an $M \times 1$ matrix representing the coordinates of the vector $|X\rangle$ in \mathcal{H}_A , and ϕ is an $N \times 1$ matrix representing the coordinates of the vector $|Y\rangle$ in \mathcal{H}_B , then the Kronecker product $\psi \otimes \phi = [\psi_i \cdot \phi]$ is an $M \cdot N \times 1$ matrix of coordinates representing $|X\rangle \otimes |Y\rangle$ in the tensor product space. However, the tensor product space also includes vectors such as $|Z\rangle = \sum \sum \gamma_{ij} \cdot (|V_i\rangle \otimes |V_j\rangle)$ such that $\gamma_{ij} \neq \psi_i \cdot \phi_j$.

A linear operator for the tensor product space maps vectors from $\mathcal{H}_A \otimes \mathcal{H}_B$ to $\mathcal{H}_A \otimes \mathcal{H}_B$. For example, suppose L is a linear operator from the \mathcal{H}_A space, and suppose M is a linear operator from the \mathcal{H}_B space. Then a linear operator for the tensor product space can be formed by the tensor product of operators $L \otimes M$, which produces the mapping

$$\begin{aligned} (L \otimes M)|Z\rangle &= (L \otimes M) \sum \sum \gamma_{ij} \cdot (|V_i\rangle \otimes |V_j\rangle) \\ &= \sum \sum \gamma_{ij} \cdot (L \otimes M)(|V_i\rangle \otimes |V_j\rangle) \\ &= \sum \sum \gamma_{ij} \cdot (L|V_i\rangle \otimes M|V_j\rangle). \end{aligned}$$

If L is the matrix representation of the operator L , and M is the matrix representation of the operator M , then the $M \cdot N \times M \cdot N$ matrix for the tensor product operator $L \otimes M$ is defined by the Kronecker product: $L \otimes M = [L_{ij} \cdot M]$. Thus, the linear transformation $(L \otimes M)(|X\rangle \otimes |Y\rangle)$ corresponds to $(L \otimes M)(\psi \otimes \phi) = L \cdot \psi \otimes M \cdot \phi$.

5.3 Quantum versus Classical Probability Theories

Quantum theory is a general theory of probability (Suppes, 1966). Like the Kolmogorov theory (Kolmogorov, 1933, 1950) it is based on axioms, but the axioms of quantum theory are different. To help introduce the basic axioms of quantum theory, we will present them side by side with the more familiar Kolmogorov axioms. Most of the applications in quantum cognition have relied on finite dimensional spaces, and so we restrict this presentation accordingly, but note that the theory can also be applied to infinite dimensional spaces. Although the space is finite, the number of dimensions can be very large (e.g., 86 billion neurons in the brain). This section provides an elementary presentation of the quantum axioms, and Section 5.2.7 includes some generalizations. See Busemeyer and Bruza (2012) and Khrennikov (2010) for a more comprehensive introduction.

5.3.1 Events

Classical probability begins by postulating a set of points or elements that define a sample space denoted as Ω . We will assume that Ω has cardinality N , and contains points $\Omega = \{\omega_1, \dots, \omega_N\}$. An event A is a subset $A \subseteq \Omega$. The collection of all events forms a Boolean algebra: if $A \subseteq \Omega$ is an event and $B \subseteq \Omega$ is an event, then the

intersection $A \cap B$ is an event that defines the conjunction; also, the union $A \cup B$ is an event that defines the disjunction, and the events satisfy the distributive rule $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Quantum theory begins by postulating a set of vectors that define a Hilbert space denoted as \mathcal{H} . We will assume that \mathcal{H} has a finite dimension N , and it is spanned by an orthonormal set of basis vectors $\mathcal{V} = \{|V_1\rangle, \dots, |V_N\rangle\}$. An event A is a subspace $\mathcal{H}_A \subset \mathcal{H}$. Changing the definition of events from subsets to subspaces changes the nature of the collection of events.

If we restrict all the events to those defined by a single basis, say $\mathcal{V} = \{|V_i\rangle, i = 1, N\}$, then the collection of events forms a Boolean algebra. Suppose event A is defined by subspace \mathcal{H}_A spanned by $\mathcal{V}_A \subset \mathcal{V}$; suppose event B is defined by subspace \mathcal{H}_B spanned by $\mathcal{V}_B \subset \mathcal{V}$. Then the conjunction, denoted $A \wedge B$, is an event spanned by $\mathcal{V}_A \cap \mathcal{V}_B$, the disjunction $A \vee B$ is an event spanned by $\mathcal{V}_A \cup \mathcal{V}_B$, and the events are distributive $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$. When a pair of events are both defined by a common basis, then they are said to be compatible.

If, however, the events are defined by subspaces that are spanned by different bases, then they are said to be incompatible, and the collection of events is not Boolean. Suppose event A is defined by subspace \mathcal{H}_A spanned by $\mathcal{V}_A \subset \mathcal{V} = \{|V_i\rangle, i = 1, N\}$; suppose event B is defined by subspace \mathcal{H}_B spanned by $\mathcal{W}_B \subset \mathcal{W} = \{|W_i\rangle, i = 1, N\}$. Then there is no meaningful way to define the conjunction or disjunction for these pairs of events, and the distributive rule does not hold (see Griffiths, 2003, chapter 4).

5.3.2 System State

In classical probability theory, we define a probability function p that maps events into probabilities. This function satisfies $p(A) \geq 0$ for all events, with $p(\emptyset) = 0$ and $p(\Omega) = 1$. If the pair of events $\{A \subseteq \Omega, B \subseteq \Omega\}$ are mutually exclusive $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$. The probability of the event “not A ,” denoted \bar{A} , equals $p(\bar{A}) = 1 - p(A)$.

In quantum probability, we define a unit length state vector $|S\rangle$ in the Hilbert space, which is then used to determine the probability of events. The probability of an event A is defined by $q(A) = \|\mathbf{P}_A|S\rangle\|^2$. If the pair of events $\{A, B\}$ are mutually exclusive, $\mathcal{H}_A \cap \mathcal{H}_B = \emptyset$, then it follows from orthogonality that $q(A \vee B) = \|\langle P_A + P_B|S\rangle\|^2 = \|\langle P_A|S\rangle\|^2 + \|\langle P_B|S\rangle\|^2 = q(A) + q(B)$. The event \bar{A} is the subspace that is the orthogonal complement to the subspace for the event A , and its probability equals $q(\bar{A}) = \|(I - P_A)|S\rangle\|^2 = 1 - q(A)$.

Note that we can express the probability of an event in terms of the inner product as follows:

$$q(A) = \|\mathbf{P}_A|S\rangle\|^2 = \left\langle S|\mathbf{P}_A^\dagger \mathbf{P}_A|S\right\rangle = \langle S|\mathbf{P}_A|S\rangle,$$

where the last step follows from the properties of projectors.

5.3.3 State Revision

According to classical probability, if an event A is observed, then a new conditional probability function is defined by the mapping $p(B|A) = \frac{p(B \cap A)}{p(A)}$. The normalizing factor in the denominator is used to guarantee that the probability assigned to the entire sample space remains equal to one. This definition forms the basis of Bayes' rule.

According to quantum probability, if an event A is observed, then the new revised state is defined by $|S_A\rangle = \frac{P_A|S\rangle}{\|P_A|S\rangle\|}$. The normalizing factor in the denominator is used to guarantee that the revised state remains unit length. The new revised state is then used (as described earlier) to compute probabilities for events. This is called Lueder's rule.

Note that we can express the conditional probability of event B given A as

$$\begin{aligned} q(B|A) &= \|P_B|S_A\rangle\|^2 \\ &= \|P_B \frac{P_A|S\rangle}{\|P_A|S\rangle\|}\|^2 \\ &= \frac{\|P_B P_A|S\rangle\|^2}{\|P_A|S\rangle\|^2}. \end{aligned}$$

5.3.4 Commutative Events

When the basis used to define events is fixed, then there is no difference between quantum and Kolmogorov theories, in the sense that the probabilities of events assigned by one can be exactly mapped into the other. In particular, for a single event, there is no need to introduce different bases.

If two projectors P_A, P_B commute $P_A P_B = P_B P_A$ then they can be decomposed in terms of the same eigenvectors, and a single basis, $\mathcal{V} = \{|V_i\rangle, i = 1, N\}$, can be used to span both events $P_A = \sum_{i \in A} |V_i\rangle \langle V_i|$ and $P_B = \sum_{i \in B} |V_i\rangle \langle V_i|$. In this case, the events are said to be compatible because they are described by the same basis. For a pair of compatible events, the joint event corresponds to the projector $P_B P_A = P_B P_A$, and the joint probability is defined by $q(A \wedge B) = \|P_B P_A|X\rangle\|^2 = \|P_A P_B|X\rangle\|^2$.

To understand more precisely how classical and quantum theories are related for commutative events, consider the following example. Suppose we have two compatible questions, labeled question A with M unique answers (e.g., how old is your daughter?) and question B with N unique answers (e.g., how is she doing in school?). All the possible answers to these two questions form $M \cdot N$ mutually exclusive and exhaustive conjunctive events $(A_i \wedge B_j), i = 1, M; j = 1, N$. Assuming that any one of these conjunctions can occur, then the cardinality of the sample space must be at least $N \cdot M$. Classical probability theory assigns a joint probability to each conjunctive event, $p(A_i \cap B_j)$, and any disjunction of events is obtained by forming the union, e.g., $A_i = \cup_j (A_i \cap B_j)$. The probability of any disjunction is obtained by summing the appropriate joint probabilities, e.g., $p(A_i) = p(\cup_j (A_i \cap B_j)) = \sum_j p(A_i \cap B_j)$.

Quantum theory represents each conjunctive event, $(A_i \wedge B_j)$, by a product of projectors $\mathbf{P}_{Ai \wedge Bj} = \mathbf{P}_{Ai} \cdot \mathbf{P}_{Bj}$, and the set of all $N \cdot M$ products produces a partition of the Hilbert space. Assuming that any one of these conjunctive events can occur, then the Hilbert space must have dimension of $N \cdot M$ or greater. Any disjunction of events is obtained by summing the appropriate conjunctive projectors, e.g., $\mathbf{P}_A = \sum_j \mathbf{P}_{Ai \wedge Bj}$. The probability of a conjunctive event equals $q(A_i \wedge B_j) = \|\mathbf{P}_{Ai} \mathbf{P}_{Bj} |S\rangle\|^2$. The probability of any other event is obtained by summing the appropriate joint probability in the same manner as classical theory, e.g., $q(A) = \|\mathbf{P}_A |S\rangle\|^2 = \left\| \sum_j \mathbf{P}_{Ai \wedge Bj} |S\rangle \right\|^2 = \sum_j q(A_i \wedge B_j)$. Suppose we use an $N \cdot M$ -dimensional Hilbert space for this case, spanned by the set of $N \cdot M$ basis vectors $\mathcal{V} = \{|V_{Ai \wedge Bj}\rangle, i = 1, M; j = 1, N\}$, and we define $\psi = [\psi_{Ai \wedge Bj}]$ as the $N \cdot M$ matrix of coordinates of $|S\rangle$ with respect to this basis. Then quantum theory can reproduce the same probabilities as classical probability theory (or vice versa) by setting $\psi_{Ai \wedge Bj} = \sqrt{p(A_i \cap B_j)}$.

Continuing with the previous example with the pair of questions A and B, classical probability theory defines two events to be independent if $p(A_i \cap B_j) = p(A_i) \cdot p(B_j)$, and otherwise these two events are dependent. The entire joint probability distribution produced by these two questions is independent if all the pairs are independent, and otherwise the distribution is dependent. An analogous definition is employed in quantum theory for compatible questions. First define a Hilbert space \mathcal{H}_A that represents question A by itself, which is spanned by basis vectors $\mathcal{V}_A = \{|V_{Ai}\rangle, i = 1, M\}$, and then define another Hilbert space \mathcal{H}_B that represents question B by itself, which is spanned by basis vectors $\mathcal{V}_B = \{|W_{Bj}\rangle, j = 1, N\}$. The complete vector space \mathcal{H} representing all of the conjunctive events and spanned by the set of $N \cdot M$ basis vectors $\mathcal{V} = \{|V_{Ai \wedge Bj}\rangle, i = 1, M; j = 1, N\}$ then can be reinterpreted as the tensor product space by setting $|V_{Ai \wedge Bj}\rangle = |V_{Ai}\rangle \otimes |V_{Bj}\rangle$. With respect to this tensor product basis,

$$|S\rangle = \sum \sum \gamma_{ij} \cdot |V_{Ai}\rangle \otimes |V_{Bj}\rangle$$

- the state is defined as *separated* if $\gamma_{ij} = \alpha_i \cdot \beta_j$,
- and otherwise it is defined as *entangled*.

Thus, an entangled state in quantum theory is analogous to a dependent joint probability distribution in classical theory. In sum, Kolmogorov and quantum probabilities will not differ for any outcomes described by the same basis. However, the theories can differ dramatically if we observe events defined by different bases. In particular, entangled states can produce marginal probabilities that cannot be reproduced by a dependent joint probability distribution when non-commuting projectors are applied to the state.

5.3.5 Non-commutative Events

The theories only differ when more than two events are involved and each event is described by a different basis. Consider the following example. Suppose you and

your opponent are playing a game in which each must choose one of N actions. Suppose that you are asked to judge the probability that you will take one action and your opponent will take another action. A classic probability model would begin by postulating an $N \times N$ joint probability distribution that describes all probabilities for the N^2 conjunctions of you and your opponent's actions. However, this assumes that these pairs of judgments do not depend on order. It may be that the judgments change depending on the order that the events are evaluated, and this signals that the events do not commute. It may be difficult to simultaneously think about your own and your opponent's preferences. Instead, it may be necessary to consider one at a time, and the order of consideration may change the answers. According to quantum theory, if the judgments depend on order, then these events are non-commutative. In this case, we can use a single N -dimensional vector space to represent both you and your opponent's perspectives. One basis for the vector space is used to evaluate your own preferences, and a different basis is used to evaluate your opponent's preferences. In this way, the quantum model captures your judgments of self versus opponent by changing the point of view used to evaluate these questions. How can one empirically determine whether events are compatible? One method is to examine the effect that question order has on the response probabilities. Incompatibility implies question order effects.

If two projectors $\mathbf{P}_A, \mathbf{P}_B$ do not commute, $\mathbf{P}_A\mathbf{P}_B \neq \mathbf{P}_B\mathbf{P}_A$, then they cannot be decomposed in terms of a single set of eigenvectors, and different bases, $\mathcal{V} = |V_i\rangle, i = 1, N$ versus $\mathcal{W} = |W_i\rangle, i = 1, N$ must be used to span both events $\mathbf{P}_A = \sum_{i \in A} |V_i\rangle \langle V_i|$ and $\mathbf{P}_B = \sum_{i \in B} |W_i\rangle \langle W_i|$. In this case, the events are said to be incompatible because they must be described by different bases. Many of the interesting aspects of quantum probability theory arise from the use of incompatible events.

According to Kolmogorov theory, the intersection event $A \cap B$ is commutative, so that $A \cap B = B \cap A$. This implies that the joint probability does not depend on order: $p(A \cap B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B) = p(B \cap A)$. However, according to quantum theory, if events A, B are incompatible then there is no such thing as an event $A \wedge B$ because the meet, \wedge , requires commutativity, but the events do not commute. Furthermore, according to quantum theory, we can only determine the probability of the sequence A and then B as follows:

$$q(A) \cdot q(B|A) = ||\mathbf{P}_B\mathbf{P}_A|\mathcal{S}\rangle|^2 \neq ||\mathbf{P}_A\mathbf{P}_B|\mathcal{S}\rangle|^2 = q(B) \cdot q(A|B).$$

The beauty of a Hilbert space is that there are many choices for the basis that one can use to describe the space. When the events are incompatible, the principle of unicity breaks down and the events cannot all be described within a single sample space (see Griffiths, 2003, chapter 27). The events spanned by the basis \mathcal{V} , which are all compatible with each other, form one sample space, and the events spanned by the basis \mathcal{W} , which are compatible with each other, form another sample space, but the events from \mathcal{V} are not compatible with the events from \mathcal{W} . In this case, from the perspective of a Kolmogorov theory, there are two stochastically unrelated

sample spaces (Dzhafarov & Kujala, 2012). Quantum theory provides a single state $|S\rangle$ that can be used to assign probabilities to both sample spaces.

5.3.6 Observables and the Uncertainty Relation

5.3.6.1 Mean and Variance of Observables

Consider the measurement of a variable X that can be assigned N real values $\{x_1, \dots, x_N\}$. If we measure the variable X , then observing the value x_i is an event. Classic probability theory defines this as a random variable X , which is a function that maps events into real numbers. The probability that X is assigned the value x_i is denoted $p(X = x_i)$, and this forms a probability distribution that sums to one: $\sum_{i=1,N} p(X = x_i) = 1$. The mean of this random variable is defined as $\mu = \sum p(X = x_i) \cdot x_i$, and the variance of this random variable is defined as $\sigma^2 = \sum p(X = x_i) \cdot (x_i - \mu)^2$.

The concept corresponding to a classical random variable in quantum theory is an observable, which is a Hermitian linear operator in the Hilbert space, defined as follows. The event x_i is represented by a projector $\mathbf{P}(x_i)$ so that the probability of observing this event equals $q(x_i) = \|\mathbf{P}(x_i)|S\rangle\|^2$. The collection of projectors, $\{\mathbf{P}(x_i), i = 1, N\}$ forms an orthogonal decomposition of the vector space: $\mathbf{P}(x_i)\mathbf{P}(x_j) = 0$ for $i \neq j$, and $\sum_i \mathbf{P}(x_i) = \mathbf{I}$. The observable for the variable X is then defined by the linear operator $\mathbf{X} = \sum_{i=1,N} \mathbf{P}(x_i) \cdot x_i$. This linear operator is Hermitian with eigenvalues equal to the values of the variable X . Given a state $|S\rangle$, the mean of the observable can be computed by the inner product

$$\mu = \langle S | \mathbf{X} | S \rangle = \sum \langle S | \mathbf{P}(x_i) | S \rangle \cdot x_i.$$

For convenience, define a new observable as $\hat{\mathbf{X}} = \mathbf{X} - \mu \cdot \mathbf{I}$. Then the variance of an observable can be computed by the inner product

$$\sigma^2 = \langle S | \hat{\mathbf{X}}^2 | S \rangle = \sum \langle S | \mathbf{P}(x_i) | S \rangle \cdot (x_i - \mu)^2.$$

5.3.6.2 Uncertainty Relation

One of the mathematical implications of quantum probability theory is the uncertainty principle. Essentially this principle states that if two observables do not commute, then a state that produces low variance on one observable must produce high variance on the other observable, so that the product of variances exceeds a non-zero boundary. This principle is not limited to physics. For example, it occurs in signal processing when trying to measure the time interval of a signal versus the frequency band of a signal (Graben & Atmanspacher, 2006). In fact, it is applicable to any system that follows the axioms of quantum probability theory, because the uncertainty principle follows from the Cauchy–Schwarz inequality (see Nielsen and Chuang, 2000, p. 89).

Suppose \mathbf{X} and \mathbf{Y} are two observables and $\hat{\mathbf{X}} = \mathbf{X} - \mu_X \mathbf{I}$ and $\hat{\mathbf{Y}} = \mathbf{Y} - \mu_Y \mathbf{I}$ and $|S\rangle$ is a state vector. The inner product $\langle S | \hat{\mathbf{X}} \hat{\mathbf{Y}} | S \rangle$ is a complex number which can be written as $\langle S | \hat{\mathbf{X}} \hat{\mathbf{Y}} | S \rangle = x + i \cdot y$. Also note that

$$\begin{aligned}
\langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} - \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle &= \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} | S \rangle - \langle S | \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \\
&= \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} | S \rangle - \langle S | \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle^* \\
&= (x + i \cdot y) - (x + i \cdot y)^* = 2 \cdot i \cdot y
\end{aligned}$$

and $\langle \psi | \hat{\mathbf{X}}\hat{\mathbf{Y}} + \hat{\mathbf{Y}}\hat{\mathbf{X}} | \psi \rangle = (x + i \cdot y) + (x + i \cdot y)^* = 2x$ so that

$$\begin{aligned}
&\left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} - \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \right|^2 + \left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} + \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \right|^2 \\
&= |2 \cdot i \cdot y|^2 + |2x|^2 = 4 \cdot (x^2 + y^2) \\
&= 4 \cdot |x + i \cdot y|^2 \\
&= 4 \cdot \left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} | S \rangle \right|^2,
\end{aligned}$$

and by the Cauchy–Schwarz inequality

$$\left| \langle S | \hat{\mathbf{X}} \cdot \hat{\mathbf{Y}} | S \rangle \right|^2 \leq \langle S | \hat{\mathbf{X}}^2 | S \rangle \cdot \langle S | \hat{\mathbf{Y}}^2 | S \rangle.$$

Combining the above two lines, we obtain the inequality

$$\begin{aligned}
\left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} - \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \right|^2 &\leq \left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} - \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \right|^2 + \left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} + \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \right|^2 \\
&= 4 \cdot \left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} | S \rangle \right|^2 \leq 4 \cdot \langle S | \hat{\mathbf{X}}^2 | S \rangle \cdot \langle S | \hat{\mathbf{Y}}^2 | S \rangle
\end{aligned}$$

and the last line implies

$$b = \frac{\left| \langle S | \hat{\mathbf{X}}\hat{\mathbf{Y}} - \hat{\mathbf{Y}}\hat{\mathbf{X}} | S \rangle \right|}{2} \leq \sqrt{\langle S | \hat{\mathbf{X}}^2 | S \rangle} \cdot \sqrt{\langle S | \hat{\mathbf{Y}}^2 | S \rangle}$$

which is the Heisenberg uncertainty relation. On the right-hand side we have the product of the standard deviations of observables \mathbf{X} and \mathbf{Y} , and this product must exceed the positive bound produced by the commutator on the left-hand side. If we prepare a person's state $|S\rangle$ for a task in such a way that the preparation decreases the variance of one observable below the bound, then this same state $|S\rangle$ must increase the variance of the other observable to exceed the bound. This provides a critical method for testing a basic principle of quantum probability theory. Psychologically, this inequality may reflect a basic limit on human capacity to attend to two incompatible observables. In addition to order effects, the uncertainty relation can be used to provide converging evidence for non-commuting observables (Atmanspacher & Roemer, 2012).

5.3.7 Generalized Measurements and Density Operators

The axioms of quantum theory can be generalized in many ways (see Nielsen & Chuang, 2000, chapter 8). One way is to generalize the definition of a measurement, and the other way is to generalize the definition of a state.

5.3.7.1 POVM: Positive Operator Valued Measurement

The general definition of measurement in quantum theory can be described as follows. Suppose we perform an experiment (e.g., ask a person a sequence of questions), and conduct a measurement to determine which one of n mutually exclusive and exhaustive sequence of answers occurs. A single sequence of answers (e.g., respond x_1 to question 1, and then x_2 to question 2, and so on) is denoted by F_i , and assume that there are n mutually exclusive and exhaustive sequences that can be observed in the experiment. Then each sequence of answers is represented by a measurement operator M_i which satisfies the completeness property: $\sum M_i^\dagger M_i = I$. The latter is required for the probabilities of answers to sum to one. In general, measurement operators that are used to represent answers do not need to be projectors, and projectors are just a special case. The probability that the measurement produces the sequence of answers F_i equals probability $\|M_i|S\rangle\|^2 = \langle S|M_i^\dagger M_i|S\rangle$. If we in fact do measure the sequence F_i , then the state is updated to $|S_{F_i}\rangle = \frac{M_i|S\rangle}{\|M_i|S\rangle\|}$. Technically, the product $E_i = M_i^\dagger M_i$ is called a positive operator value measurement (POVM), and M_i is a measurement operator that produces this POVM. However, more than one measurement operator can produce the same POVM.

For example, consider the measurement of question A with answers (yes,no) followed by the measurement of question B with answers (yes,no). Suppose P_{Ay} is the projector representing the answer yes to question A, and $P_{An} = I - P_{Ay}$ is the orthogonal projector for the answer no to question A. Suppose P_{By} is the projector representing the answer yes to question B, and $P_{Bn} = I - P_{By}$ is the orthogonal projector for the answer no to question B. Also assume that these two projectors do not commute, $P_{Ay}P_{By} \neq P_{By}P_{Ay}$. Then there are four possible sequences of answers (A_y, B_y) , (A_y, B_n) , (A_n, B_y) , (A_n, B_n) . We can assign four measurement operators to these four answers as follows: $M_{yy} = P_{By}P_{Ay}$, $M_{ny} = P_{Bn}P_{Ay}$, $M_{yn} = P_{By}P_{An}$, $M_{nn} = P_{Bn}P_{An}$. These measurement operators are *not* projectors; however, they do satisfy the completeness property.

5.3.7.2 Density Operators and Mixed States

In Section 5.2.2, the state of the system was represented by a unit length vector $|S\rangle$ in the Hilbert space. Technically, this is called a pure state. A pure state can be changed into what is called a *density* operator by forming the outer product: $\rho = |S\rangle\langle S|$. Note ρ is the projector for state $|S\rangle$, so that $Tr[\rho^2] = Tr[\rho] = Tr[|S\rangle\langle S|] = \langle S|S\rangle = 1$. If we define the state as the density operator, then the probability of a sequence of answers F_i , represented by measurement operator M_i , is given by

$$q(F_i) = Tr[M_i\rho M_i^\dagger] = Tr[M_i|S\rangle\langle S|M_i^\dagger] = \langle S|M_i^\dagger M_i|S\rangle.$$

The state after the measurement is then updated to $\rho_{Fi} = \frac{M_i\rho M_i^\dagger}{q(F_i)}$.

Pure states are special cases of a density operator that satisfy $Tr[\rho^2] = 1$. Quantum theory allows for more general types of density operators, called mixed

states, that only require that ρ be a positive operator and that $Tr[\rho] = 1$. A mixed state occurs when we do not know the exact pure state, and instead we form a probability distribution over pure states: $\rho = \sum p_i \cdot \rho_i$, with $\rho_i = |S_i\rangle\langle S_i|$, $p_i \geq 0$ and $\sum p_i = 1$. In psychology, this mixed state is useful when applying the theory to an ensemble of participants (Camparo, 2013). However, the decomposition of a mixed state is not unique. Mixed states only satisfy the inequality $Tr[\rho^2] \leq 1$. The probability of a sequence of answers F , represented by measurement operator M , is given by $q(F) = Tr[M\rho M^\dagger] = \sum p_i Tr[M\rho_i M^\dagger]$.

One might question whether the principles for computing probabilities in quantum theory are arbitrary. This turns out not to be true. A famous theorem by Gleason (1957) proves that this is *the* unique method for assigning an additive probability measure to events defined as subspaces of a Hilbert space with dimension greater than 3.

5.3.8 Psychological Applications

Three examples are used to illustrate the application of quantum theory to psychology. One concerns probability judgment errors, the second concerns order effects on attitude judgments, and the third describes a violation of rational decision-making. The power of quantum theory is to use the same principles to account for widely different phenomena.

5.3.9 Conjunction Fallacy

First we apply the theory to the well-known conjunction fallacy (Tversky & Kahneman, 1983). For some specially designed judgment problems, individuals systematically tend to judge the probability of a conjunction ($A \wedge B$) to be greater than one of the single events (e.g., B assuming B is judged to be less likely than A). One popular example is the “Linda” problem in which participants are initially told a story about a liberal and well-educated woman named Linda, and then they are asked to choose which of two events is more likely: (a) Linda is a feminist bank teller, versus (b) Linda is a bank teller. The majority of participants tend to choose the “feminist and bank teller” event as more likely than “bank teller.”

We apply quantum theory to this finding as follows (Busemeyer *et al.*, 2011). See Aerts (2009) for an alternative account using quantum theory. First, we assume that the “Linda” story determines the initial state $|S\rangle$. The event “feminist” is represented by a projector P_F and the event “not a feminist” is then defined by $P_{\bar{F}} = I - P_F$. The event “bank teller” is represented by the projector P_B . The probability of the event “feminist and bank teller” is determined by a sequence of projections as follows: $q(F, B) = \|P_B P_F |S\rangle\|^2$. The probability of the event “bank teller” is determined by a single projection $q(B) = \|P_B |S\rangle\|^2$. However, the latter can be expanded as follows

$$\begin{aligned} q(B) &= \|P_B |S\rangle\|^2 \\ &= \|P_B I |S\rangle\|^2 \end{aligned}$$

$$\begin{aligned}
&= \|\mathbf{P}_B (\mathbf{P}_F + \mathbf{P}_{\bar{F}}) |S\rangle\|^2 \\
&= \|\mathbf{P}_B \mathbf{P}_F |S\rangle + \mathbf{P}_B \mathbf{P}_{\bar{F}} |S\rangle\|^2 \\
&= \|\mathbf{P}_B \mathbf{P}_F |S\rangle\|^2 + \|\mathbf{P}_B \mathbf{P}_{\bar{F}} |S\rangle\|^2 + Int,
\end{aligned}$$

where $Int = \langle S | \mathbf{P}_{\bar{F}} \mathbf{P}_B \mathbf{P}_F | S \rangle + \langle S | \mathbf{P}_F \mathbf{P}_B \mathbf{P}_{\bar{F}} | S \rangle$ is called the interference term, which can be positive or negative or zero. Note that this interference term is zero if the projectors commute. In order to account for the empirical fact that $p(F, B) > p(B)$, we require that $\|\mathbf{P}_B \mathbf{P}_F |S\rangle\|^2 > \|\mathbf{P}_B \mathbf{P}_F |S\rangle\|^2 + \|\mathbf{P}_B \mathbf{P}_{\bar{F}} |S\rangle\|^2 + Int$, which implies that $\|\mathbf{P}_B \mathbf{P}_{\bar{F}} |S\rangle\|^2 + Int < 0$ to account for the conjunction fallacy. Of course, the latter is *post hoc*, but now that we have committed to the assumption that the projectors do not commute, we must make other *a priori* predictions for probability judgments (Busemeyer *et al.*, 2011). One, in particular, is that we must predict that the evaluation of the conjunction depends on the order of this evaluation. In fact, past research has found order effects with these types of comparisons (Stolarz-Fantino, *et al.*, 2003).

5.3.10 Question Order Effects

Question order effects are very common in attitude and survey research. An interesting example is the result of a 1997 poll in the USA concerning public opinions on the characters of Bill Clinton and Al Gore (Moore, 2002). Half of the 1,002 nationally sampled respondents were asked: “Do you generally think Bill Clinton is honest and trustworthy?” Subsequently, they were asked the same question about Al Gore. The other half of respondents answered exactly the same questions, but in the opposite order. The results exhibited a striking order effect: the proportion saying “yes” to both questions significantly increased when Gore was judged first.

Quantum theory predicts question order effects whenever the projectors do not commute. However, more than just predicting order effects, quantum theory predicts a particular pattern of order effects. Suppose that participants are asked to answer (yes, no) to two different questions (A, B) in opposite orders. Define $p(Ay, Bn)$, for example, as the probability of saying yes to question A and then no to question B when question A was asked first; likewise, define $p(Bn, Ay)$, for example, as the probability of saying no to question B and then yes to question A when question B was asked first; similar definitions are used for the remaining sequences of answers. Suppose \mathbf{P}_{Ay} is the projector corresponding to the answer yes to question A and $\mathbf{P}_{An} = \mathbf{I} - \mathbf{P}_{Ay}$ is the projector for no to question A. Likewise, suppose \mathbf{P}_{By} is the projector corresponding to the answer yes to question B and $\mathbf{P}_{Bn} = \mathbf{I} - \mathbf{P}_{By}$ is the projector for no to question B. According to quantum theory,

$$\begin{aligned}
q(Ay, Bn) &= \|\mathbf{P}_{Bn} \mathbf{P}_{Ay} |S\rangle\|^2 \\
q(Bn, Ay) &= \|\mathbf{P}_{Ay} \mathbf{P}_{Bn} |S\rangle\|^2.
\end{aligned}$$

If we use the more general form based on density operators, quantum theory predicts

$$\begin{aligned} q(Ay, Bn) &= \text{Tr} [\mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho \mathbf{P}_{Ay} \mathbf{P}_{Bn}] \\ q(Bn, Ay) &= \text{Tr} [\mathbf{P}_{Ay} \mathbf{P}_{Bn} \rho \mathbf{P}_{Bn} \mathbf{P}_{Ay}]. \end{aligned}$$

Based on this basic model for order effects, we can derive the following *a priori* and parameter-free quantitative prediction called the QQ equality (Wang & Busemeyer, 2013; Wang *et al.*, 2014):

$$(q(Ay, Bn) + q(An, By)) - (q(By, An) + q(Bn, By)) = 0.$$

This was originally proved in Wang and Busemeyer (2013), but a simpler proof was recently produced by James Yearsley (personal communication), which is presented below. Before presenting the proof, recall that $[\mathbf{L}, \mathbf{M}] = (\mathbf{LM} - \mathbf{ML})$ is the commutator for linear operators \mathbf{L}, \mathbf{M} . Also note that the following properties hold for the projectors: $[\mathbf{P}_{Ay}, \mathbf{P}_{By}] = [\mathbf{P}_{Ay}, \mathbf{I} - \mathbf{P}_{Bn}] = -[\mathbf{P}_{Ay}, \mathbf{P}_{Bn}] = [\mathbf{P}_{Bn}, \mathbf{P}_{Ay}]$. Then the QQ equality is derived as follows. Obviously,

$$[\mathbf{P}_{Ay}, \mathbf{P}_{By}] - [\mathbf{P}_{Ay}, \mathbf{P}_{By}] = 0$$

which implies

$$\mathbf{P}_{Ay} [\mathbf{P}_{Ay}, \mathbf{P}_{By}] + \mathbf{P}_{An} [\mathbf{P}_{Ay}, \mathbf{P}_{By}] - [\mathbf{P}_{Ay}, \mathbf{P}_{By}] \mathbf{P}_{By} - [\mathbf{P}_{Ay}, \mathbf{P}_{By}] \mathbf{P}_{Bn} = 0.$$

Using the previously mentioned property of projectors, we make the following substitutions for the commutators:

$$\mathbf{P}_{Ay} [\mathbf{P}_{Bn}, \mathbf{P}_{Ay}] + \mathbf{P}_{An} [\mathbf{P}_{By}, \mathbf{P}_{An}] - [\mathbf{P}_{By}, \mathbf{P}_{An}] \mathbf{P}_{By} - [\mathbf{P}_{Bn}, \mathbf{P}_{Ay}] \mathbf{P}_{Bn} = 0.$$

Assuming as before that the measurements are projectors, then expanding the commutators produces

$$(\mathbf{P}_{Ay} \mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho + \mathbf{P}_{An} \mathbf{P}_{By} \mathbf{P}_{An} \rho) - (\mathbf{P}_{By} \mathbf{P}_{An} \mathbf{P}_{By} \rho + \mathbf{P}_{Bn} \mathbf{P}_{Ay} \mathbf{P}_{Bn} \rho) = 0.$$

Multiplying both sides by an arbitrary density operator produces

$$(\mathbf{P}_{Ay} \mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho + \mathbf{P}_{An} \mathbf{P}_{By} \mathbf{P}_{An} \rho) - (\mathbf{P}_{By} \mathbf{P}_{An} \mathbf{P}_{By} \rho + \mathbf{P}_{Bn} \mathbf{P}_{Ay} \mathbf{P}_{Bn} \rho) = 0.$$

Taking the trace (and recalling that the trace is a linear operator) produces

$$\begin{aligned} &(\text{Tr} [\mathbf{P}_{Ay} \mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho] + \text{Tr} [\mathbf{P}_{An} \mathbf{P}_{By} \mathbf{P}_{An} \rho]) \\ &- (\text{Tr} [\mathbf{P}_{By} \mathbf{P}_{An} \mathbf{P}_{By} \rho] + \text{Tr} [\mathbf{P}_{Bn} \mathbf{P}_{Ay} \mathbf{P}_{Bn} \rho]) = 0. \end{aligned}$$

Once again, assuming that the measurements are projectors, then $\text{Tr} [\mathbf{P}_{Ay} \mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho] = \text{Tr} [\mathbf{P}_{Ay} \mathbf{P}_{Bn} \mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho]$. The cyclic property of trace implies that

$$\text{Tr} [(\mathbf{P}_{Ay} \mathbf{P}_{Bn})(\mathbf{P}_{Bn} \mathbf{P}_{Ay}) \rho] = \text{Tr} [\mathbf{P}_{Bn} \mathbf{P}_{Ay} \rho \mathbf{P}_{Ay} \mathbf{P}_{Bn}] = q(A_y, B_n).$$

This last result finally leads to the QQ equality

$$(q(A_y, B_n) + q(A_n, B_y)) - (q(B_y, A_n) + q(B_n, A_y)) = 0.$$

It is worth noting that this proof requires two assumptions: (1) questions are asked back to back with no intervening information; (2) the measurements are represented by projectors. The proof does not hold for more general POVM type of measurements.

The QQ equality is an *a priori* and parameter-free prediction that provides a very strong empirical test of the quantum model for order effects. Recently, research has shown that it was statistically supported across 70 national field experiments that examined question order effects using a wide range of topics (Wang *et al.*, 2014).

5.3.11 Disjunction Effect

The third application is to a finding called the disjunction effect (Tversky & Shafir, 1992). The disjunction effect was discovered in the process of testing a rational principle of decision theory proposed by Savage (1954) called “the sure thing principle”: if you prefer action A over B under state of the world X, and you also prefer action A over B under the complementary state of the world $\sim X$ (“not X”), then you should prefer action A over B even if the state of the world is unknown. The earliest studies used a two-stage gamble, which is the same gamble (equal chance to win \$200 or lose \$100) played twice. The first play was obligatory, and the player had to decide whether to play again on the second round. When asked to imagine winning the first play, a majority (69%) chose to play the second round; when asked to imagine losing the first round, again a majority (59%) chose to play the second round; but when the first play was considered unknown, a minority (36%) chose to play the second round. Tversky and Shafir argued that this result violates the sure thing principle of decision theory.

Quantum theory accounts for the disjunction effect as follows (Pothos & Busemeyer, 2009). The information about the first-stage play of the gamble places the decision-maker into one of three states: if the first play of the gamble is known to win, then the state conditioned on this fact is set equal to $|S_W\rangle$; if the first play of the gamble is known to lose, then the state conditioned on this fact is set equal to $|S_L\rangle$; if the first play of the two-stage gamble is unknown, then the person is superposed between the two known states so that the unknown state equals $|S_U\rangle = \frac{1}{\sqrt{2}} \cdot |S_W\rangle + \frac{1}{\sqrt{2}} \cdot |S_L\rangle$. The decision to play the gamble is represented by a projector \mathbf{P}_G . Pothos and Busemeyer (2009) designed a specific form for this projector based on the utilities of the gamble. However, whatever the form used for this projector, it follows that the probability to play the gamble on the second stage equals

$$\begin{aligned} q(\text{Play} \mid \text{win}) &= \|\mathbf{P}_G |S_W\rangle\|^2 \\ q(\text{Play} \mid \text{lose}) &= \|\mathbf{P}_G |S_L\rangle\|^2 \\ q(\text{Play} \mid \text{unkown}) &= \|\mathbf{P}_G |S_U\rangle\|^2 \\ &= \left\| \frac{1}{\sqrt{2}} \cdot \mathbf{P}_G |S_W\rangle + \frac{1}{\sqrt{2}} \cdot \mathbf{P}_G |S_L\rangle \right\|^2 \\ &= \frac{1}{2} \cdot \|\mathbf{P}_G |S_W\rangle\|^2 + \frac{1}{2} \cdot \|\mathbf{P}_G |S_L\rangle\|^2 + \text{Int}, \end{aligned}$$

where $Int = \frac{1}{2} \cdot (\langle S_W | \mathbf{P}_G | S_L \rangle + \langle S_L | \mathbf{P}_G | S_W \rangle)$. To account for the results, we require that $Int < 0$ and sufficiently negative to lower $q(Play | unkown)$ well below either $q(Play | win)$ or $q(Play | lose)$.

Critics might argue that quantum models' ability to account for all these findings relies on these models being more complex than traditional models. This criticism has been directly addressed by performing a rigorous quantitative comparison of quantum vs. previously successful traditional decision models (e.g., prospect theory proposed by Tversky & Shafir, 1992) using a Bayesian model comparison method (Busemeyer, Wang, & Shiffrin, 2015). The Bayesian model comparison method evaluates models with respect to their accuracy, parsimony, and robustness. The Bayes factor strongly favored the quantum model over the traditional decision models, suggesting the quantum model is more accurate and robust and not necessarily more complex.

5.4 Markov versus Quantum Dynamics

The primary classical analysis of probabilistic dynamical systems is provided by Markov processes. Quantum theory also provides a general theory of probabilistic dynamical systems. As we will see below, there are close analogies between quantum and Markov dynamics, but the two also differ in fundamental ways. Below we present a parallel development of Markov and quantum systems. Once again, this presentation is primarily limited to finite N -dimensional systems (although N can be quite large). However, both systems can be extended to infinite dimensional systems.

5.4.1 System State

Both quantum and Markov types of systems are described in terms of states that the system enters at different time points. For example, in a signal detection type of experiment, the decision-maker accumulates confidence regarding the presence or absence of a signal by collecting evidence from samples of information across time. In this case, the states are represented by different levels of confidence that the signal is absent versus present.

For both Markov and quantum types of systems, suppose that there is a set of $N = 2n + 1$ basis states (e.g., basic levels of confidence) for the system,

$$\Omega = \{|S_{-n}\rangle, |S_0\rangle, \dots, |S_{+n}\rangle\}.$$

For example, $|S_{-n}\rangle$ may represent a state of high confidence that signal is not present, $|S_0\rangle$ may represent a state of uncertainty, and $|S_{+n}\rangle$ may represent a state of high confidence that the signal is present. For both types of systems, the set of basis states Ω can be interpreted as a set of basis vectors that span an N -dimensional vector space. For the Markov system, the vector space is defined on a field of real numbers; but for the quantum system, the vector space is defined on a field of complex numbers.

According to Markov theory, the system is located in exactly one basis state at any moment in time, and the system moves from one basis state to another across time to form a trajectory. Define $|S(t)\rangle \in \Omega$ as the state of the Markov system at time t . We (the theorists) do not usually know the location of the state of the system at any moment in time. Therefore, we assign a probability that the system is located in a basis state at each time as follows: $p_i(t) = Pr [|S(t)\rangle = |S_i\rangle]$, with $p_i(t)$ a real number, $0 \leq p_i(t) \leq 1$ and $\sum p_i(t) = 1$. With respect to the Ω basis, each basis vector $|S_j\rangle$ can be represented by an $N \times 1$ column matrix $S_j = [\delta(i,j)]$ with $\delta(i,j) = 1$ for row $i = j$ and $\delta(i,j) = 0$ for rows $i \neq j$. Then the *mixed* state of the Markov system at time t is defined as the $N \times 1$ column matrix

$$\mathbf{p}(t) = \sum_i p_i(t) \cdot S_i,$$

which is called the probability distribution over the basis states. It is convenient for summing probabilities to define a $1 \times N$ row matrix containing all ones denoted $\mathbf{I} = [1, 1, \dots, 1]$. Then $\mathbf{p}(t) \in \text{span}(\Omega)$ under the positivity restriction that $p_i \geq 0$ and the linear restriction that $\mathbf{I} \cdot \mathbf{p}(t) = 1$ for all t .

According to quantum theory, the system is *not* necessarily located in exactly one basis state at any moment in time, so that the system does not have a well-defined trajectory across time. Instead, the state is superposed over the basis states at each moment in time. The state of the quantum system is defined by the superposition

$$|S(t)\rangle = \sum \psi_i(t) \cdot |S_i\rangle,$$

where ψ_i is a complex number and $\sum |\psi_i(t)|^2 = 1$. Once again, with respect to the Ω basis, each basis vector $|S_j\rangle$ can be represented by an $N \times 1$ column matrix $S_j = [\delta(i,j)]$. Using this representation, the quantum state can be represented by the $N \times 1$ column matrix

$$\boldsymbol{\psi}(t) = \sum \psi_i(t) \cdot S_i,$$

which is called the amplitude distribution over basis states. Then $\boldsymbol{\psi}(t) \in \text{span}(\Omega)$ under unit length restriction that $\|\boldsymbol{\psi}(t)\|^2 = 1$ for all t .

5.4.2 Transitions between States

Both quantum and Markov types of systems describe the dynamic evolution of probabilities across states of the system. Markov systems evolve the probability distribution across time, whereas quantum systems evolve the amplitude distribution across time.

For Markov systems, the probability of transitioning to state $|S_i\rangle$ at time t_2 from state $|S_j\rangle$ at time t_1 is denoted as $T_{ij}(t_1, t_2)$. These transition probabilities form an $N \times N$ transition matrix $T(t_1, t_2) = [T_{ij}(t_1, t_2)]$. The evolution of the probability distribution over states from time t_1 to time t_2 is determined by the linear transformation

$$\mathbf{p}(t_2) = T(t_2, t_1) \cdot \mathbf{p}(t_1).$$

The transition matrix $T(t_2, t_1)$ must be a stochastic matrix to guarantee that $\mathbf{p}(t_2)$ remains a probability distribution. A stochastic matrix is one with non-negative entries that satisfy $\sum_i T_{ij}(t_2, t_1) = 1$. Repeated application of the transition matrix produces the result $\mathbf{p}(t_3) = T(t_3, t_2) \cdot \mathbf{p}(t_2) = T(t_3, t_2) \cdot T(t_2, t_1) \cdot \mathbf{p}(t_1)$, which implies that the transition matrix satisfies the Chapman–Kolmogorov equation

$$T(t_3, t_1) = T(t_3, t_2) \cdot T(t_2, t_1).$$

The Markov system is said to be stationary if

$$T(t_1 + t, t_1) = T(t_2 + t, t_2)$$

for all t_1, t_2 , in which case the transition probabilities only depend on the time interval: $T(t_2, t_1) = T(t_2 - t_1)$. Stationary Markov systems satisfy the semi group property: $T(t + s) = T(t) \cdot T(s)$. From the semi group property, it follows that the transition matrix satisfies the Kolmogorov foward equation

$$\frac{d}{dt} T(t) = K \cdot T(t),$$

where K is the $N \times N$ intensity matrix (Bhattacharya & Waymire, 1990). Post multiplying both sides by the probability distribution provides the equation for the evolution of probabilities

$$\frac{d}{dt} \mathbf{p}(t) = K \cdot \mathbf{p}(t).$$

The intensity matrix K must satisfy $K_{ij} \geq 0$ for $i \neq j$ and $\sum_i K_{ij} = 0$ to guarantee that $T(t)$ remains a stochastic matrix. The solution to the Kolmogorov forward equation is given by the matrix exponential

$$\begin{aligned} T(t) &= \exp(t \cdot K) \\ \mathbf{p}(t) &= \exp(t \cdot K) \cdot \mathbf{p}(0). \end{aligned}$$

For quantum systems, the amplitude of transiting to state $|S_i\rangle$ at time t_2 from state $|S_j\rangle$ at time t_1 is denoted as $U_{ij}(t_1, t_2)$. These transition amplitudes form an $N \times N$ unitary matrix $U(t_1, t_2) = [U_{ij}(t_1, t_2)]$. The evolution of the amplitude distribution over states from time t_1 to time t_2 is determined by the linear transformation

$$\psi(t_2) = U(t_2, t_1) \cdot \psi(t_1).$$

The matrix $U(t_2, t_1)$ must be unitary, $U^\dagger U = I$, to guarantee that $\psi(t_2)$ remains an amplitude distribution. We can derive a transition matrix $T(t_2, t_1)$ from the unitary matrix by taking the squared magnitudes of the transition amplitudes $T_{ij}(t_2, t_1) = |U_{ij}(t_2, t_1)|^2$. The transition matrix formed in this way is doubly stochastic: it has non-negative entries that satisfy $\sum_i T_{ij}(t_2, t_1) = 1$ and $\sum_j T_{ij}(t_2, t_1) = 1$. Repeated application of the unitary matrix produces the result $\psi(t_3) = U(t_3, t_2) \cdot \psi(t_2) = U(t_3, t_2) \cdot U(t_2, t_1) \cdot \psi(t_1)$, which implies that the unitary matrix satisfies the product rule

$$U(t_3, t_1) = U(t_3, t_2) \cdot U(t_2, t_1).$$

The quantum system is said to be stationary if

$$U(t_1 + t, t_1) = U(t_2 + t, t_2)$$

for all t_1, t_2 , in which case the transition amplitudes only depend on the time interval: $U(t_2, t_1) = U(t_2 - t_1)$. Stationary quantum systems satisfy the semi group property: $U(t + s) = U(t) \cdot U(s)$. From the semi group property, it follows that the unitary matrix satisfies the Schrödinger equation

$$\frac{d}{dt}U(t) = -i \cdot H \cdot U(t),$$

where H is a $N \times N$ Hamiltonian matrix (Hughes, 1989). Post multiplying each side by the amplitude distribution provides the equation for the evolution of the amplitudes

$$\frac{d}{dt}\psi(t) = -i \cdot H \cdot \psi(t).$$

The $N \times N$ Hamiltonian matrix H must be the Hermitian matrix, $H = H^\dagger$, to guarantee that $U(t)$ remains a unitary matrix. The solution to the Schrödinger equation is given by the matrix exponential

$$U(t) = \exp(-i \cdot t \cdot H),$$

$$\psi(t) = \exp(-i \cdot t \cdot H) \cdot \psi(0).$$

Note that if $\psi(0)$ is an eigenvector of H , then there is no change in state (except for a change in global phase) and the state is an equilibrium state.

5.4.3 Measurements

We can take a measurement of the state of either a Markov or a quantum system at any point in time. Suppose that we apply a measurement that produces a value from the set $\{R_1, \dots, R_j, \dots, R_n\}$ with $1 < n \leq N$. Define $R(t)$ as the observed response from the measurement at time t . Assume that each value obtained from the measurement corresponds to a projector $P_j = \sum_{i \in R_j} S_i \cdot S_i^\dagger$, such that $P_i \cdot P_j = 0$ for $i \neq j$ and $\sum_j P_j = I$, $j = 1, N$. Thus P_j simply selects the coordinates that we assume are assigned to the value R_j .

If we measure the Markov system at time t , then the probability of observing value R_j equals

$$p(R(t) = R_j) = \mathbf{1} \cdot P_j \cdot \mathbf{p}(t).$$

If we observe the value R_j at time t , the state of the Markov system is updated based on this observation:

$$\mathbf{p}(t|R(t) = R_j) = \frac{P_j \mathbf{p}(t)}{p(R(t) = R_j)}.$$

The updated distribution, conditioned on observing this observation, sums to one. Suppose we take repeated measurements of the Markov system at time points $\{t_1, t_2, \dots, t_m\}$. Then the probability of the sequence of observed values equals

$$\begin{aligned} p(R(t_m), \dots, R(t_2), R(t_1)) &= \\ &\mathbf{I} \cdot P_{R(t_m)} T(t_m, t_{m-1}) \cdots P_{t_2} \cdot T(t_2, t_1) \cdot P_{R(t_1)} \cdot T(t_1, t_0) \cdot \mathbf{p}(t_0). \end{aligned}$$

If we measure the quantum system at time t , then the probability of observing value R_j equals

$$p(R(t) = R_j) = \|P_j \cdot \psi(t)\|^2.$$

If we observe the value R_j at time t , the state of the quantum system is updated based on this observation:

$$\psi(t|R(t) = R_j) = \frac{P_j \psi(t)}{\sqrt{p(R(t) = R_j)}}.$$

The updated distribution, conditioned on observing this observation, has unit length. Suppose we take repeated measurements of the quantum system at time points $\{t_1, t_2, \dots, t_m\}$. Then the probability of the sequence of observed values equals

$$\begin{aligned} q(R(t_m), \dots, R(t_2), R(t_1)) &= \\ &= \|P_{R(t_m)} U(t_m, t_{m-1}) \cdots P_{t_2} \cdot U(t_2, t_1) \cdot P_{R(t_1)} \cdot U(t_1, t_0) \cdot \psi(t_0)\|^2. \end{aligned}$$

5.4.4 Generalized Measurements and Density Operators

5.4.4.1 Hidden Markov Models and Quantum Noise Models

It is not necessary to represent measurements by projectors. Instead, we can assign a more general measurement operator M_j to each value R_j .

For the Markov model, we can assign a measurement operator to response R_j so that $p(R(t) = R_j) = \mathbf{I} \cdot M_j \cdot \mathbf{p}(t)$. We require that the collection satisfies the completeness property $\sum_j M_j = I$ so that the response probabilities sum to $\mathbf{I} \cdot \sum_j M_j \cdot \mathbf{p}(t) = 1$. For example, suppose $M_i = \text{diag}[m_{i1}, m_{i2}, \dots, m_{iN}]$ for $j = 1, n$ with the requirement that $1 \geq m_{ij} \geq 0$ and $\sum_i m_{ij} = 1$ for $j = 1, N$. Then the observations of response R_j could have been obtained from any of the N states, and the state remains hidden after the measurement. However, we can infer the probability distribution over the states from the measurement using the updating equation: $\mathbf{p}(t|R(t) = R_j) = \frac{M_j \mathbf{p}(t)}{p(R(t) = R_j)}$.

For the quantum model, we can also assign a measurement operator to response R_j so that $q(R(t) = R_j) = \|M_j \cdot \psi(t)\|^2$. We require that the collection satisfies the completeness property $\sum_j M_j^\dagger M_j = I$ so that the response probabilities sum to $\sum_j \psi(t)^\dagger M_j^\dagger M_j \psi(t) = 1$. For example, suppose $M_i = \text{diag}[\sqrt{m_{i1}}, \sqrt{m_{i2}}, \dots, \sqrt{m_{iN}}]$ for $j = 1, n$ with the requirement that $1 \geq m_{ij} \geq 0$ and $\sum_i m_{ij} = 1$ for $j = 1, N$. These types of measurements are considered

to be noisy quantum measurements. Once again, we can infer the probability distribution over the states from the measurement using the updating equation:

$$p(t|R(t) = R_j) = \frac{M_j p(t)}{p(R(t)=R_j)}.$$

5.4.4.2 Density Matrices

The density matrix corresponding to the state ψ equals $\rho(t) = \psi(t) \cdot \psi(t)^\dagger$. The state evolves across time according to the Schrödinger equation and so the density evolves according to

$$\begin{aligned} \frac{d}{dt}\rho(t) &= \frac{d}{dt}(\psi(t) \cdot \psi(t)^\dagger) \\ &= \frac{d}{dt}(\psi(t)) \cdot \psi(t)^\dagger + \psi(t) \cdot \frac{d}{dt}(\psi(t)^\dagger) \\ &= (-i \cdot H) \psi(t) \cdot \psi(t)^\dagger + \psi(t) \cdot \psi(t)^\dagger (i \cdot H) \\ &= -i \cdot (H \cdot \rho(t) - \rho(t) \cdot H) \\ &= -i \cdot [H, \rho(t)]. \end{aligned}$$

The solution to this differential equation then equals

$$\begin{aligned} \rho(t) &= \psi(t) \cdot \psi(t)^\dagger \\ &= U(t) \cdot \rho(0) \cdot U^\dagger(t). \end{aligned}$$

Note that if H commutes with $\rho(0)$, then there is no evolution and the state is an equilibrium state located at one of the eigenvectors of H . Mixed states are simply linear combinations of the pure states, and so the same dynamic equations apply to the mixed states.

5.4.5 Psychological Applications

5.4.5.1 Interference of Choice on Later Confidence

Kvam *et al.* (2015) compared the ability of Markov and quantum random walk models to predict the effect of making a choice on subsequent ratings of confidence in a motion detection task. Participants were shown a set of dots on screen, the majority of which were appearing and disappearing in random locations. However, a subset of the dots were coherently moving to the left or to the right – participants made decisions and judgments regarding the direction of motion for this subset of the dots.

There were two main conditions. In the choice condition, participants were asked to report whether they believed the dots were moving left at an initial time point t_1 , and then rate their confidence from 0 (certain of left motion) to 100 (certain of right motion) at a second time point t_2 . In the no-choice condition, they simply made a motor response at t_1 and then rated their confidence at t_2 . Evidence accumulation can therefore be divided into two stages, $t_0 - t_1$ and $t_1 - t_2$ (although naturally there will be some time taken out to make a response at t_1). There were a total of nine participants and each participant experienced both conditions on

different trial blocks. The coherence (percentage of dots moving in one direction) was also manipulated as well as the time interval between t_1 and t_2 , but here we focus on the choice versus no-choice manipulation.

Both models are used to predict the choice at time point t_1 (for the choice condition) and the confidence rating at time point t_2 . However, we are primarily concerned with the marginal distribution of confidence ratings produced at time point t_2 (pooled across choices made at time point t_1 for the choice condition). As we prove later, the two models make qualitatively different predictions concerning the difference between conditions: the Markov model predicts no effect of the choice on the marginal distribution of confidence ratings, but the quantum model predicts an interference effect of the choice on the marginal distribution of confidence ratings. Before presenting the proof, we first present more details about the implementation of each model.

Each model uses $N = 103$ basis states $\{|S_{-1}\rangle, \dots, |S_{101}\rangle\}$. States 0 to 100 map directly onto the 100 confidence levels, and states -1 and 101 serve as reflecting boundaries that correspond to confidence levels 0 and 100, respectively. As in the signal detection model above, we assume unbiased initial beliefs, so initial states are set to be uniformly distributed around state $|S_{50}\rangle$ with width w . For the Markov model, this yields an initial mixed state vector $\mathbf{p}(0) = [p_j(0)]$, where $p_j(0) = \frac{1}{w}$ for $(50 - \frac{w}{2}) \leq j \leq (50 + \frac{w}{2})$. For the quantum model, it yields an initial superposition state vector $\psi(0) = [\psi_j(0)]$, where $\psi_j(0) = \frac{1}{\sqrt{w}}$ for $(50 - \frac{w}{2}) \leq j \leq (50 + \frac{w}{2})$.

We also define matrices to describe the state dynamics corresponding to evidence accumulation during the task. In the Markov model, this is given by the $N \times N$ intensity matrix $K = [k_{m,n}]$, with $k_{m,m} = -\sigma^2$, $k_{m-1,m} = \frac{\sigma^2-\delta}{2}$, $k_{m+1,m} = \frac{\sigma^2+\delta}{2}$. In the quantum model, it is given by the $N \times N$ Hamiltonian $H = [h_{m,n}]$, with $h_{m,m} = \frac{\delta \cdot m}{103}$ and $h_{m,m+1} = h_{m,m-1} = \sigma^2$. For simplicity, we assume that the dots are moving to the right, so that participants are sampling rightward motion information at a rate δ and randomly sampling leftward or rightward information at a rate σ^2 . However, these could be easily adjusted to give transition matrices corresponding to leftward dot motion.

The intensity and Hamiltonian matrices allow us to construct dynamic operators that take the initial state $\mathbf{p}(0)$ or $\psi(0)$ in order to give their positions at time t_1 . For the Markov model, the first stage transition matrix is $T(t_1) = \exp(t_1 \cdot K)$, and the state at time t_1 is $\mathbf{p}(t_1) = T(t_1) \cdot \mathbf{p}(0)$. In the no-choice condition, this state remains. However, when a decision is made in the Markov model, it informs the observer about the state of the system. For example, if a person responds “right,” then they are not in any of states $\{|S_{-1}\rangle, \dots, |S_{49}\rangle\}$ because these all correspond to “left” response states, assuming an unbiased choice criterion. Therefore, when a choice is made, we are left with two conditional distributions, $\mathbf{p}(t_1|'right')$ and $\mathbf{p}(t_1|'left')$. In order to obtain these distributions, we must specify the response operators M_R and M_L that map the state onto the “left” and “right” responses, respectively.

$$M_L = 0.5 \cdot |S_{50}\rangle + \sum_{j=-1}^{49} |S_j\rangle$$

$$M_R = 0.5 \cdot |S_{50}\rangle + \sum_{j=51}^{101} |S_j\rangle$$

The application of these measurement operators yields a state whose entries do not sum to one. Therefore, the new conditional state must be normalized by the probability of obtaining that state, which is given by the summing the entries lying on the corresponding side of the scale.

$$\begin{aligned} p(t_1|'right') &= \frac{M_R \cdot p(t_1)}{0.5 \cdot p_{50}(t_1) + \sum_{j=51}^{101} p_j(t_1)} \\ p(t_1|'left') &= \frac{M_L \cdot p(t_1)}{0.5 \cdot p_{50}(t_1) + \sum_{j=-1}^{49} p_j(t_1)} \end{aligned}$$

The same intensity matrix Q is used to compute the transition matrix for the second stage of information processing, $T(t_2, t_1) = \exp([t_2 - t_1] \cdot \gamma \cdot K)$. The additional factor γ is used to specify the amount of information processed per unit time in the second stage relative to the first stage (Kvam *et al.*, 2015; Yu, Pleskac, & Zeigenfuse, 2015). The marginal distribution of confidence following a choice is given by taking each of the final states following a “left” or “right” choice ($p(t_2|'left') = T(t_2, t_1) \cdot p(t_1|'left')$ and $p(t_2|'right') = T(t_2, t_1) \cdot p(t_1|'right')$), multiplying each one by the probability of obtaining the corresponding response at t_1 , and adding the two together.

Finally, the probability of obtaining a particular confidence response y is given by applying the response operator M_y , which is an $N \times N$ matrix with zeros everywhere except for ones along the diagonal in the rows corresponding to confidence level y . These are then summed together using operator I , a $1 \times N$ matrix of ones.

$$\begin{aligned} Pr(y|choice) &= Pr('right' \wedge conf = y) + Pr('left' \wedge conf = y) \\ &= I \cdot [M_y \cdot T(t_2, t_1) \cdot M_R \cdot T(t_1) \cdot p(0) + M_y \cdot T(t_2, t_1) \cdot M_L \cdot T(t_1) \cdot p(0)] \\ &= I \cdot M_y \cdot T(t_2, t_1) \cdot [M_R \cdot T(t_1) \cdot p(0) + M_L \cdot T(t_1) \cdot p(0)] \\ &= I \cdot M_y \cdot T(t_2, t_1) \cdot I \cdot T(t_1) \cdot p(0) \\ &= Pr(y|no\ choice). \end{aligned}$$

As we can see, the Markov model predicts the same marginal probability of responding at each confidence level between the choice and no-choice conditions described above. Because the system is located in one basis state at any given time, the dynamic Markov model does not naturally predict any effect of choice at t_1 on confidence at t_2 .

The measurement and dynamic operators in the quantum model, however, lead to a different prediction. Using the Hamiltonian specified above, we obtain the unitary matrix $U(t_1) = \exp(-i \cdot t_1 \cdot H)$ that describes the change in the system based on the first stage of information processing. This gives the state at time t_1 , $\psi(t_1) = U(t_1) \cdot \psi(0)$. As in the Markov model, the no-choice state remains the same. For the choice condition, we compute two conditional states for “right” and “left” choices using two response operators, P_R and P_L .

$$P_L = \sqrt{0.5} \cdot |S_{50}\rangle + \sum_{j=-1}^{49} |S_j\rangle$$

$$P_R = \sqrt{0.5} \cdot |S_{50}\rangle + \sum_{j=51}^{101} |S_j\rangle.$$

The probability of a “right” response at t_1 is the squared length of the projection $||P_R \cdot \psi(t_1)||^2$, and the probability of a “left” response is $||P_L \cdot \psi(t_1)||^2$. The conditional states are computed by projecting the states onto the corresponding halves of the scale, and normalized by the (non-squared) length of this vector so that the new states are again unit length.

$$\psi(t_1)'right' = \frac{P_R \cdot \psi(t_1)}{||P_R \cdot \psi(t_1)||}$$

$$\psi(t_1)'left' = \frac{P_L \cdot \psi(t_1)}{||P_L \cdot \psi(t_1)||}$$

The unitary matrix describing the second stage of information processing, applied to each of the three (no-choice, choose left, choose right) states at t_1 , is $U(t_2, t_1) = \exp(-i \cdot [t_2 - t_1] \cdot \gamma \cdot H)$. As in the Markov model, γ indicates the amount of information that participants are processing per unit time in the second stage relative to the first stage.

The same confidence response operator M_y can be used to map the evidence states onto confidence level y , and the $1 \times N$ matrix of ones I is again used to sum across rows of the state vectors.

$$\begin{aligned} Pr(y|choice) &= Pr('right', conf = y) + Pr('left', conf = y) \\ &= I \cdot ||M_y \cdot U(t_2, t_1) \cdot P_R \cdot U(t_1) \cdot \psi(0)||^2 + ||M_y \cdot U(t_2, t_1) \cdot P_L \cdot U(t_1) \cdot \psi(0)||^2 \\ &\neq I \cdot ||M_y \cdot U(t_2, t_1) \cdot P_R \cdot U(t_1) \cdot \psi(0) + M_y \cdot U(t_2) \cdot P_L \cdot U(t_1) \cdot \psi(0)||^2 \\ &= I \cdot ||M_y \cdot U(t_2, t_1) \cdot I \cdot U(t_1) \cdot \psi(0)||^2 \\ &= Pr(y|no\ choice). \end{aligned}$$

The reason for the inequality in line 4 is an interference term that arises when it is evaluated, similar to the one shown in Section 5.3.9. Because of this term, there can be an *interference effect* where making a decision at t_1 produces a different distribution over states at t_2 relative to the condition where no decision is made.

The experimental results by Kvam *et al.* (2015) found that confidence judgments in the no-choice condition were more extreme than those in the choice condition, indicating that choice had produced an interference effect that showed up in subsequent confidence ratings. In addition, the authors computed a Bayes factor between the models, yielding an estimate of the relative credibility of the two models given the available data. They found strong support for the quantum model over the Markov model on the group level as well as the majority of individual participants. This suggests that the quantum model provides a better account of the empirical data on top of its ability to predict interference in this design.

5.5 Concluding Comments

In this chapter, we have introduced quantum probability as a framework from which to construct probabilistic models of cognitive processes. Models using these principles have been used to examine new and existing questions in decision-making, judgments, memory, concept combinations, perception, and evidence accumulation. While we did not delve into specific models for all of these areas of application, we hope to have provided the mathematical and conceptual basis for why and how one might implement quantum models of such processes.

Much of the motivation for constructing these models is inherently tied to the underlying mathematical principles of quantum theory. Quantum probability is built upon a logic of vector spaces, treating events as subspaces, states as complex superpositions, dynamics as unitary transformations, and measurements as projection operators. Taking this approach to modeling leads to unique predictions, including non-commutativity in transformations (order effects), interference effects between sequential measurements, the uncertainty principle for incompatible bases, and context dependency in observed outcomes.

A critical question, tied into empirical evidence for the quantum predictions, is whether this framework is more appropriate for modeling cognition than the classical probability framework on which most current models are based. The classical approach treats events as sets and subsets of points, states as probability distributions, dynamics as transitions between states, and measurement as simply an inspection of the location of the state. It posits the existence of a joint distribution over all possible combinations of events, leading to predictions that diverge from the quantum framework. In particular, we looked at each theory's prediction for the effect of measurement on the system, (non-)commutative events, and the uncertainty relations between pairs of observable events.

When applied to cognition, these predictions can be tested empirically. We examined applications to the conjunction fallacy, question orders, the disjunction effect, and dynamic problems like the evolution of confidence ratings over time and the effect of measurements (choice) on subsequent confidence responses.

5.5.1 Open Questions and Problems

While we have explored some examples of the empirical work which has taken a quantum approach to cognition, there are many remaining issues and unexamined consequences which we have not covered. Two important features of this approach are that

1. an entirely different set of underlying axioms are used to specify the likelihood of events and how they are realized; and
2. there are a wide range of problems in physics and computer science (especially information theory) that have psychological analogues.

These two factors yield a wealth of new, empirically testable predictions for quantum cognitive models. For example, various authors have proposed tests of temporal Bell inequalities (Atmanspacher & Filk, 2010), the triangle inequality in similarity (Tversky & Gati, 1982; Pothos *et al.*, 2013), non-local correlations between entangled cognitive systems (Vedral, 2008), complementarity (Graben & Atmanspacher, 2008; Primas, 2007), and distributivity (Birkhoff & von Neumann, 1936; Wang *et al.*, 2013).

The outcomes of these tests aim primarily at falsifying classical probability principles. However, empirical tests may also potentially challenge the quantum framework. A Hilbert space representation requires that the law of double stochasticity holds, so that all transformations $U = [u_{ij}]$ meet the condition that the squared length of the entries in each column and each row sum to one, $\sum_j ||u_{ij}||^2 = 1$ and $\sum_i ||u_{ij}||^2 = 1$. In addition, projectors P_A and P_B in a Hilbert space must obey the law of reciprocity, so that $|\langle P_A | P_B \rangle| = |\langle P_B | P_A \rangle|$. A consequence of this property has been formalized and tested, called the *Q-Q equality* (Wang & Busemeyer, 2013; Wang *et al.*, 2014), but this does not imply that it cannot be violated.

Of course, simply establishing a mathematical principle does not necessarily lead one to how it could be tested empirically. One of the greatest challenges that modelers face is constructing experimental paradigms that are capable of diagnosing violations of classical or quantum principles. On this issue, we have provided some examples of paradigms that have been previously used, but these may often need to be modified or foregone depending on the problem.

5.5.2 Extensions and Alternatives to Quantum Probability

All of the examples that we have provided, and many of the core concepts, are based on ideal quantum models which use state vectors and noiseless evolution and measurement operators. We went into some detail in Section 5.3.7 about how these can be generalized using density matrices, noise operators, and POVMs. These concepts have been used to construct quantum models of response times (Fuss & Navarro, 2008, 2013), but a great deal remains to be done to examine how

noise affects the particular models we have presented and what methods are most appropriate for constructing quantum models in different domains.

In addition, while classical and quantum probability are the most prominent foundations for modeling, these are not the only possible theories of probability. Indeed, a number of contextual probability theories (Dzhafarov & Kujala, 2013) as well as ones based on different event algebras or belief functions (Narens, 2007) produce interesting results and can explain a number of psychological phenomena. The necessity of these generalizations would be elucidated by theoretically expounding and empirically examining their assumptions, but such work is also likely to shed light on the usefulness of both classical and quantum approaches.

References

- Aerts, D. (2009). Quantum structure in cognition. *Journal of Mathematical Psychology*, 53(5), 314–348.
- Aerts, D., & Aerts, S. (1994). Applications of quantum statistics in psychological studies of decision processes. *Foundations of Science*, 1, 85–97.
- Aerts, D., & Gabora, L. (2005). A theory of concepts and their combinations II: A Hilbert space representation. *Kybernetes*, 34, 192–221.
- Atmanspacher, H., & Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. *Journal of Mathematical Psychology*, 54, 314–321.
- Atmanspacher, H., Filk, T., & Romer, H. (2004). Quantum zero features of bistable perception. *Biological Cybernetics*, 90, 33–40.
- Atmanspacher, H., & Roemer, H. (2012). Order effects in sequential measurements of non-commuting psychological observables. *Journal of Mathematical Psychology*, 56(4), 274–280.
- Bhattacharya, R. N., & Waymire, E. C. (1990). *Stochastic processes with applications*. New York, NY: Wiley.
- Birkhoff, G., & von Neumann, J. (1936). The logic of quantum mechanics. *Annals of Mathematics*, 37, 823–843.
- Blutner, R. (2009). Concepts and bounded rationality: An application of Niestegge's approach to conditional quantum probabilities. In L. Acardi *et al.* (eds), *Foundations of probability and physics – 5*, Vol. 1101, pp. 302–310. New York, NY: Springer.
- Bordley, R. F., & Kadane, J. B. (1999). Experiment-dependent priors in psychology. *Theory and Decision*, 47(3), 213–227.
- Brainerd, C. J., Wang, Z., & Reyna, V. (2013). Superposition of episodic memories: Overdistribution and quantum models. *Topics in Cognitive Science*, 5(4), 773–799.
- Bruza, P., Kitto, K., Nelson, D., & McEvoy, C. (2009). Is there something quantum-like in the human mental lexicon? *Journal of Mathematical Psychology*, 53, 362–377.
- Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge: Cambridge University Press.
- Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, 118(2), 193–218.

- Busemeyer, J. R., Wang, Z., & Shiffrin, R. M. (2015). Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency. *Decision*, 2, 1–12.
- Busemeyer, J. R., Wang, Z., & Townsend, J. (2006). Quantum dynamics of human decision making. *Journal of Mathematical Psychology*, 50(3), 220–241.
- Camparo, J. (2013). A geometrical approach to the ordinal data of Likert scaling and attitude measurements: The density matrix in psychology. *Journal of Mathematical Psychology*, 57(1), 29–42.
- Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., & Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems and Information Dynamics*, 16, 1–17.
- Dirac, P. A. M. (1958). *The principles of quantum mechanics*. Oxford: Oxford University Press.
- Dzhafarov, E., & Kujala, J. V. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology*, 56, 54–63.
- Dzhafarov, E., & Kujala, J. V. (2013). All-possible-couplings approach to measuring probabilistic context. *PLoS ONE*, 8(5), e61712.
- Franco, R. (2009). Quantum amplitude amplification algorithm: An explanation of availability bias. In P. Bruza, D. Sofge, W. Lawless, K. van Rijsbergen, & M. Klusch (eds.), *Quantum Interaction: Lecture notes in computer science*, Vol. 5494 (pp. 84–96). New York, NY: Springer.
- Fuss, I. G., & Navarro, D. J. (2008). Partially coherent quantum models for human two choice decisions. In P. Bruza, W. Lawless, K. van Rijsbergen, D. Sofge, B. Coecke, & S. Clark (eds.), *Proceedings of the second quantum interaction symposium*, (pp. 75–82). London: College Publications.
- Fuss, I. G., & Navarro, D. J. (2013). Open parallel cooperative and competitive decision processes: A potential provenance for quantum probability decision models. *Topics in Cognitive Science*, 5(4), 818–843.
- Gleason, A. M. (1957). Measures on the closed subspaces of a Hilbert space. *Journal of Mathematical Mechanics*, 6, 885–893.
- Graben, P. B., & Atmanspacher, H. (2006). Complementarity in classical dynamical systems. *Foundations of Physics*, 36, 291–306.
- Graben, P. B., & Atmanspacher, H. (2008). Extending the philosophical significance of the idea of complementarity. In H. Atmanspacher & H. Primas (eds.), *Recasting reality: Wolfgang Pauli's philosophical ideas and contemporary science* (pp. 99–113). Berlin: Springer.
- Griffiths, R. B. (2003). *Consistent quantum theory*. Cambridge: Cambridge University Press.
- Hameroff, S. R. (1998). Quantum computation in brain microtubules? The Penrose–Hameroff ‘Orch OR’ model of consciousness. *Philosophical Transactions Royal Society London (A)*, 356, 1869–1896.
- Hughes, R. I. G. (1989). *The structure and interpretation of quantum mechanics*. Cambridge, MA: Harvard University Press.
- Ivancevic, V. G., & Ivancevic, T. T. (2010). *Quantum neural computation*. New York, NY: Springer.
- Khrennikov, A. Y. (2010). *Ubiquitous quantum structure: From psychology to finance*. New York, NY: Springer.

- Khrennikov, A. Y., & Haven, E. (2009). Quantum mechanics and violations of the sure-thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology*, 53(5), 378–388.
- Kolmogorov, A. N. (1933/1950). *Foundations of the theory of probability*. New York, NY: Chelsea Publishing Co.
- Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences* 112(34), 10645–10650.
- Lambert-Mogiliansky, A., Zamir, S., & Zwirn, H. (2009). Type indeterminacy: A model of the KT (Kahneman–Tversky)-man. *Journal of Mathematical Psychology*, 53(5), 349–361.
- La Mura, P. (2009). Projective expected utility. *Journal of Mathematical Psychology*, 53(5), 408–414.
- Moore, D. W. (2002). Measuring new types of question-order effects. *Public Opinion Quarterly*, 66, 80–91.
- Narens, L. (2007). *Theories of probability: An examination of logical and qualitative foundations*. Hackensack, NJ: World Scientific Publishing Company.
- Nielsen, M., & Chuang, I. (2000). *Quantum computation and quantum information*. Cambridge: Cambridge University Press.
- Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of ‘rational’ decision-making. *Proceedings of the Royal Society, B*, 276(1665), 2171–2178.
- Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36(3), 255–274.
- Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review*, 120(3), 679–696.
- Primas, H. (2007). Non-Boolean descriptions for mind-matter problems. *Mind & Matter*, 5(1), 7–44.
- Savage, L. J. (1954). *The foundations of statistics*. New York, NY: John Wiley & Sons.
- Stolarz-Fantino, S., Fantino, E., Zizzo, D. J., & Wen, J. (2003). The conjunction effect: New evidence for robustness. *American Journal of Psychology*, 116(1), 15–34.
- Suppes, P. (1966). The probabilistic argument for a nonclassical logic in quantum mechanics. *Philosophy of Science*, 33, 14–21.
- Trueblood, J. S., & Busemeyer, J. R. (2010). A quantum probability account for order effects on inference. *Cognitive Science*, 35, 1518–1552.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological review*, 89(2), 123–154.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunctive fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science*, 3, 305–309.
- Vedral, V. (2008). Quantifying entanglement in macroscopic systems. *Nature*, 453(7198), 1004–1007.
- von Neumann, J. (1932/1955). *Mathematical foundations of quantum mechanics*. Princeton, NJ: Princeton University Press.

- Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, 5(4), 689–710.
- Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, 5(4), 672–688.
- Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences of the USA*, 111(26), 9431–9436.
- Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, 144(2), 489.
- Yukalov, V. I., & Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision*, 70, 283–328.

6 Computational Cognitive Neuroscience

F. Gregory Ashby

6.1	Introduction	224
6.1.1	A Brief History	224
6.1.2	Organization of the Chapter	226
6.2	Advantages of CCN Modeling	226
6.2.1	Testing Against Many Different Data Types	227
6.2.2	Model Inflexibility	227
6.2.3	Model Convergence	228
6.2.4	Ability to Unite Seemingly Disparate Fields	228
6.3	CCN Modeling Principles	229
6.3.1	The Neuroscience Ideal	230
6.3.2	The Simplicity Heuristic	230
6.3.3	The Set-in-Stone Ideal	231
6.4	Models of Single Spiking Neurons	231
6.4.1	The Leaky Integrate-and-Fire Model	232
6.4.2	The Quadratic Integrate-and-Fire Model	234
6.4.3	The Izhikevich Model	235
6.4.4	Modeling Synaptic Delays	238
6.4.5	Noise	240
6.5	Firing-rate Models	241
6.6	Learning	244
6.6.1	Synaptic Plasticity	244
6.6.2	Models of Learning in the Striatum and Cortex	246
6.6.2.1	Discrete-time Models of Learning at Synapses that Lack Fast DA Reuptake	247
6.6.2.2	Discrete-time Models of Learning at Synapses with Fast DA Reuptake	248
6.6.2.3	Modeling DA Release	250
6.6.2.4	Continuous-time Models of Hebbian Learning	252
6.7	Testing CCN Models	254
6.7.1	Single-unit Recording Data	254
6.7.2	Behavioral Data	255
6.7.3	fMRI Data	257
6.7.4	TMS Data	261
6.7.5	Pharmacological and Neuropsychological Patient Data	262

6.8 Parameter Estimation and Model Evaluation	262
6.9 Conclusions	264
Acknowledgment	264
References	264

6.1 Introduction

Cognitive neuroscience was born in the 1990s amid a technological explosion that produced powerful new methods for non-invasively studying the human brain, including functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation (TMS). These exciting new technologies revolutionized the scientific study of the mind, giving unprecedented observability into the neural processes that mediate human thought and action. With the new data came a growing need for new kinds of theories that could simultaneously account for the behavioral data that are the bread and butter of traditional mathematical psychology as well as the brain-related measures coming from the new technologies. Computational cognitive neuroscience (CCN) was created to fill this void.

CCN evolved from computational neuroscience on one side and connectionism, neural network theory, and machine learning on the other. Like computational neuroscience, CCN strives for neurobiological accuracy, and like connectionism, a major goal is to account for behavior. However, CCN is unique because most computational neuroscience models make no attempt to account for behavior and most connectionist models make no attempt to be biologically detailed. The biological detail included in CCN models adds many more constraints on the modeling process than more traditional approaches. As a result, two researchers independently modeling the same behavior are more likely to converge on highly similar models with this new approach, and for this reason the resulting models should have a permanence that is unusual with older approaches. Not surprisingly, a growing number of researchers are now pursuing the endeavor of CCN modeling (e.g., Anderson *et al.*, 2008; Ashby & Hélie, 2011; Frank, 2005; Hartley *et al.*, 2000; O'Reilly *et al.*, 2012).

6.1.1 A Brief History

The field of computational neuroscience became popular with Hodgkin and Huxley's (1952) Nobel Prize-winning efforts to model the generation of action potentials in the squid giant axon. Like the Hodgkin and Huxley model, most computational neuroscience models include only a single neuron. A common approach, called compartment modeling, models a neuron's axons and dendrites as cylinders and the soma as a sphere. Next, partial differential equations that describe the propagation of action potentials are written for each of these compartments. A standard application is to model the results of patch-clamp experiments in which current is injected into the neuron at some location and then the intracellular voltage is measured at a variety of locations on the cell. Some compartment models are

extremely accurate, but highly complex. For example, some single-neuron models have hundreds or even thousands of compartments (e.g., Bhalla & Bower, 1993; Segev, Fleshman, & Burke, 1989). Historically, computational neuroscience models have almost never tried to account for behavior. In most cases, such a goal is precluded by the complexity of the single-neuron models that are used.

Neural network theory originated in the work of McCulloch and Pitts (1943). However, because the goal quickly became to model or at least simulate behavior, neural network theory diverged from computational neuroscience with the work of Newell, Shaw, and Simon (1958) and Rosenblatt (1958). At that time, computing power was too limited and there simply was not enough known about the neural basis of behavior to support a research program that tried to model behavior in a biologically accurate way. Thus the fields of artificial intelligence and the more modern related field of machine learning place almost all emphasis on behavior and almost none on neuroscience.

Connectionism (Rumelhart & McClelland, 1986) and modern neural network theory (e.g., Haykin, 2009) take an intermediate approach in the sense that biologically plausible properties are often seen as advantages, although they rarely are requirements. Connectionist models have some features in common with the brain – including distributed representation, continuous flow, and the modeling of memory as changes in synaptic strengths. Even so, almost all connectionist models include many features that are now known to be incompatible with brain function. For example, there is generally no attempt to identify units in connectionist models with specific brain regions, and even when there is, there is little attempt to model inputs and outputs to these regions in a biologically accurate way. Similarly, units in connectionist models typically do not behave like real neurons, and the learning algorithms that are used often have little biological plausibility (e.g., backpropagation).

Each of these fields makes important contributions. Computational neuroscience provides a formal framework to test theories of biophysics. Artificial intelligence and machine learning allow engineers to construct machines and algorithms that exhibit intelligent behavior. Connectionism allows psychologists to construct neurally inspired models of behaviors that are so complex or poorly understood that it would be premature to attempt to build more biologically detailed models. CCN is not meant to supplant these older approaches, but rather to fill a new void that was created by the cognitive neuroscience revolution. There are now many behaviors that are understood well enough at the neural level to permit biologically detailed mathematical modeling. CCN was born in an attempt to exploit these new data.

Because of this motivation, it is not surprising that the field of CCN began shortly after the onset of the cognitive neuroscience revolution during the 1990s. The first break with existing approaches came with attempts to associate nodes in fairly traditional connectionist or neural network models with specific brain regions. This trend toward increased biological detail continued with more biologically plausible learning algorithms, and more realistic models of the individual units (e.g., Ashby *et al.*, 1998; Cohen, Braver, & O'Reilly, 1996; Cohen & Servan-Schreiber, 1992; McClelland, McNaughton, & O'Reilly, 1995). Simultaneously,

there were also attempts to formulate general modeling principles of this new approach (Ashby & Hélie, 2011; Ashby & Valentin, 2007; O'Reilly, 1998; O'Reilly *et al.*, 2012).

6.1.2 Organization of the Chapter

This chapter is organized as follows. Section 6.2 describes some advantages of the CCN approach. Section 6.3 describes some of the CCN principles that guide model development and model testing. Section 6.4 describes some common approaches used in CCN models in which each unit models a single spiking neuron. Section 6.5 describes an alternative approach that models the firing rates of large populations of neurons. Section 6.6 describes models of learning that are based on the neuroscience literature on long-lasting synaptic plasticity. CCN models make predictions about neural activations in each brain region included in the model. To test the model against empirical data, some modeling interface is typically needed that converts the predicted neural activations to the dependent measure that defines the data. Section 6.7 describes these interfaces and other relevant issues for applications to single-unit recording data, behavioral data [i.e., accuracy and response time (RT)], fMRI BOLD responses, TMS data, and pharmacological and neuropsychological patient data. Section 6.8 discusses the problem of parameter estimation and model evaluation, and Section 6.9 closes with some general comments and conclusions.

6.2 Advantages of CCN Modeling

Most of the advantages of CCN over more traditional, purely cognitive mathematical modeling are due to the many neuroscience-derived constraints that CCN models must satisfy – constraints that drastically reduce the space of compatible models. Architecturally, the construction of CCN models is sharply constrained by neuroanatomy and basic neuroscience results. For example, if the model includes a cortical region and the striatum then neuroanatomy tells us that the only possible CCN model of these two regions is one where the cortex sends an excitatory projection to the striatum and there is no direct return projection from the striatum to the cortex. Restrictions such as these are in sharp contrast with methods used to construct traditional cognitive or connectionist models, where any architecture is allowed, and alternative models are evaluated almost solely on the basis of goodness-of-fit to available RT and accuracy data.

CCN models are constrained not only in architecture, but also in process. For example, neuroanatomy specifies whether each connection is excitatory or inhibitory, and single-unit recording data serve to constrain the dynamics of individual units in the model (e.g., whether the units have a high or low tonic firing rate, whether they fire in bursts or at a steady rate). Furthermore, neuroscience data also sharply constrain how learning and memory are modeled. For example, the evidence is good that dopamine (DA) mediates feedback-driven synaptic plasticity in the striatum (e.g., Doya, 2007; Schultz, 2002). When DA neurons fire, DA

is released approximately uniformly throughout large regions of the striatum, and as a result, any CCN model that includes striatal-mediated learning must assume global rather than local learning rules.

The more constrained CCN modeling process confers a number of specific advantages to CCN models. The remainder of this section describes the most important of these.

6.2.1 Testing Against Many Different Data Types

Whereas cognitive models can generally be tested only against RT and accuracy data, CCN models can be tested against many different kinds of dependent measures. Theoretically, this should include virtually any dependent measure between behavior at the highest level and single-unit recording data at the lowest level. So, for example, the same CCN model could be tested against RTs, accuracies, single-neuron recording data, fMRI BOLD responses, and EEG recordings. In addition, CCN provides a principled method for modeling the effects of various interventions that affect neural processes, including TMS, neuropharmacological treatment, and neurosurgical procedures such as ablation or deep-brain stimulation. Similar approaches could be used to account for behavioral deficits that are associated with various neuropsychological conditions (e.g., Parkinson's disease, Huntington's disease, anterograde amnesia). Requiring successful models to simultaneously account for more different kinds of data necessarily increases model identifiability.

6.2.2 Model Inflexibility

CCN models often include many unknown constants (or parameters) that must be set (or estimated) during the model-fitting process. A traditional cognitive-based mathematical model with the same number of free parameters would be so mathematically flexible that it would be difficult to falsify on the basis of goodness-of-fit alone. This principle is immortalized in the following famous quote attributed to John von Neumann (by Enrico Fermi): "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk" (Dyson, 2004).

Von Neumann's quote is most apt for David Marr's (1982) highest level of mathematical modeling – what he called the computational level (and what is often referred to in mathematical psychology as the descriptive level). When a parameter is added to a computational-level model, its sole purpose is almost always to fit more data. Therefore, it is almost always true that every new parameter greatly increases mathematical flexibility. However, as one moves down the Marr hierarchy, the quote becomes less and less relevant. At Marr's algorithmic level (called process models in mathematical psychology), new parameters are added to model some new (e.g., psychological) process. Modeling a new process will generally increase mathematical flexibility, but one new process could add several new parameters that work together. This trend continues at Marr's lowest, implementational level where the goal is to model the hardware that implements

the algorithms. Now new parameters are added to model new structures and in many cases, a new process might require several new structures, each of which requires several new parameters to model.

For these reasons, implementational-level models generally have many parameters but are nevertheless mathematically inflexible. CCN models are at Marr's implementational level, and as expected, they tend to be very inflexible. The inflexibility is built in via the architectural and process constraints supplied by the relevant neuroscience literature. For example, consider a CCN model that includes cortical and striatal units. The equations describing each unit will be characterized by a number of free parameters (for details, see Section 6.4) and there will be other parameters that describe the strength of the cortical–striatal synapses. However, because the projection from cortex to striatum is excitatory and one-way, changing the values of any of these parameters can only have a very limited effect on the behavior of the model. For every unit, there are two sets of parameter settings: those that allow the model to fire spikes and those for which the model never fires spikes. The latter settings are disallowed because they fail to produce neuron-like behavior. Any parameter combination in the former group of settings will cause similar behavior in the model – namely, any condition that causes cortical units to increase their firing rate will also cause striatal units to increase their firing rate. This is the only data profile that the model can produce, regardless of how many free parameters it contains, and regardless of the numerical values of those parameters.

This inflexibility is a strength of CCN models. If the results are qualitatively incompatible with the CCN predictions, then the architectural and process assumptions of the model are almost always incorrect. It is rarely the case that the mispredictions are caused by poor parameter estimation. As a result, invalid models are quickly rejected, hastening the scientific process.

6.2.3 Model Convergence

Another major advantage of the CCN approach is that two researchers independently modeling the same behavior are more likely to converge on highly similar models, and this convergence should cause the resulting models to have a permanence that is unusual with more traditional approaches. For example, the evidence is overwhelming that the hippocampus plays an important role in episodic memory consolidation. So any CCN model of episodic memory is likely to include a component that models some region(s) in the hippocampus. Since the neuroanatomy of the hippocampus is well-understood, independently constructed CCN models of episodic memory should therefore include some highly similar components.

6.2.4 Ability to Unite Seemingly Disparate Fields

CCN modeling can sometimes uncover relationships among seemingly unrelated behaviors. This is especially likely when independent modeling efforts converge on common brain regions. For example, cognitive neuroscience models of

information-integration category learning and implicit sequence learning independently identified similar cortical–striatal circuits (e.g., Ashby *et al.*, 1998; Grafton, Hazeltine, & Ivry, 1995). This suggested that these two seemingly disparate behaviors might share some previously unknown deep functional similarity. Several studies explored this possibility. First, Willingham, Wells, Farrell, and Stemwedel (2000) showed that implicit motor sequence production is disrupted when the response key locations are switched, but not when the hands used to depress the keys are switched. Several studies subsequently showed that this same pattern of results holds for information-integration categorization (Ashby, Ell, & Waldron, 2003; Maddox, Bohil, & Ing, 2004; Maddox *et al.*, 2010). Without linking categorization and sequence learning through their hypothesized underlying neural circuits, this dependence of information-integration categorization on response location learning would have been much more difficult to discover.

CCN models can also sometimes establish links between models that do not share any common brain regions. Researchers who build and test CCN models are typically interested in one sub-process more than others. For example, researchers interested in visual perception might couple a detailed model of some regions in the visual cortex with an oversimplified model of response selection that produces a stylized output, whereas researchers interested in motor performance might couple a detailed model of the primary motor cortex with an oversimplified model of visual perception that produces a stylized input to the motor module.

If different researchers each build valid CCN models of their module of interest, then it should be fairly straightforward to create a new model by linking the two separate CCN models together, and this new model should be consistent with all the behavioral and neuroscience data that are consistent with either model alone. Furthermore, by relying on neuroscience data, CCN modeling can even instruct us on *how* to connect the models together (by looking at which regions project to which; e.g., Cantwell *et al.*, 2017). This ability to combine models of different processes is not trivial; even if each of two models work in their given domain, there is no guarantee that the combined model will behave as desired.

6.3 CCN Modeling Principles

In traditional cognitive-based mathematical modeling of behavior, the overriding criterion for establishing the validity of a model is goodness-of-fit to the behavioral data (usually penalized for model complexity; see, e.g., Pitt, Myung, & Zhang, 2002). Unfortunately, there are many examples where models making very different cognitive assumptions provide approximately equal levels of goodness-of-fit, so in many cognitive domains there are many competing mathematical models that make very different cognitive assumptions. In most cases, there is not much that can be done to resolve these problems. One solution is to invoke the *principle of correspondent change* (Townsend & Ashby, 1983), which

states that the correct model should account for behavioral changes across experimental conditions by only changing the values of parameters that correspond to the independent variables that were varied to create the various conditions. So, for example, if the different conditions are identical except for the brightness of the stimulus, then to fit the resulting data the correct model should only need to change the value of sensory parameters (so decision parameters should be invariant across conditions). Although the principle of correspondent change is rarely invoked, when used it can prove effective (e.g., Van Zandt, Colonius, & Proctor, 2000).

The principle of correspondent change is an attempt to add extra constraints to computational- and algorithmic-level models. The constraints that define CCN can also be formalized as a set of principles (Ashby & Hélie, 2011; Meeter, Jehee, & Murre, 2007; O'Reilly, 1998). This section describes three of the five principles that were proposed as constraints on CCN modeling by Ashby and Hélie (2011). For more discussion, or an alternative list of principles, see Ashby and Hélie (2011), Meeter *et al.* (2007), or O'Reilly (1998).

6.3.1 The Neuroscience Ideal

A CCN model should not make any assumptions that are known to contradict the current neuroscience literature.

This principle formalizes the CCN goal of building models that are constrained by existing neuroscience data. Note, however, that the neuroscience ideal does not say that a CCN model must be compatible with all existing neuroscience data. A model is an abstraction, and therefore is almost always incomplete. The brain is immensely complex, and every CCN model must omit much of this complexity. The neuroscience ideal weighs these errors of omission much less heavily than errors of commission (Meeter *et al.*, 2007). For instance, it is common in CCN modeling to omit connections between some brain areas that are known to be connected. This is done to keep the model simple and to focus on other connections that the model assumes are functionally important. A good fit suggests the missing connections might not be functionally important to the phenomenon under study. On the other hand, the neuroscience ideal disallows connections between brain areas that are known to not be connected or to create an inhibitory connection between two regions when it is known that the real projection is excitatory.

6.3.2 The Simplicity Heuristic

Because of the great complexity of the human brain, every CCN modeling project must decide on an appropriate level of reductionism. The Simplicity Heuristic is a guide to solving this problem.

No extra neuroscientific detail should be added to the model unless there are data to test this component of the model or past research has shown that the neuroscientific detail is a major contributor to the explanation.

This heuristic is an application of Occam's razor. It is especially important because unlike cognitive models, with CCN models there will almost always be many extra neuroscientific details that one could add to an existing model. Doing so will increase the complexity of the model, the number of free parameters, and the computing time required for fitting. Unless there are data to test these extra components, it will be impossible to know whether these extra details were modeled correctly, and to what extent these untested details contributed to the model's success.

6.3.3 The Set-in-Stone Ideal

Once set, the architecture of the network and the models of each individual unit should remain fixed throughout all applications.

This could be considered a corollary to the principle of correspondent change. Connections between brain regions do not change from task to task, nor does the qualitative nature via which a neuron responds to input. Thus, the model's analogues of these features should also not change when the empirical application changes. This ideal greatly reduces the mathematical flexibility of CCN models. For example, although a CCN model will initially have many unknown constants, most of these will be set by single-unit recording data and then, by the set-in-stone ideal, they will remain invariant across all applications of the model.

6.4 Models of Single Spiking Neurons

The units that comprise the neural network in a CCN model should include more biological detail than the units used in traditional neural network theory (e.g., Haykin, 2009) or connectionism (Rumelhart & McClelland, 1986). CCN models follow two different general approaches to this problem. One approach builds networks from units that mimic spiking neurons, and another builds networks that model the firing rates of populations of neurons in different brain regions. This section describes spiking-neuron models and the next section describes firing-rate models.

The spiking-neuron models used in CCN originate in the classic Hodgkin–Huxley model (1952), which is a set of four coupled differential equations. One describes fast changes in intracellular voltage and three describe slow changes in various ion concentrations (i.e., Na^+ , K^+ , and Cl^-). The model correctly accounts for action potentials (both the upstroke and downstroke), the refractory period, and subthreshold depolarizations that fail to produce a spike. From a CCN perspective, the model has several disadvantages. First, it was created to model voltage changes in the squid giant axon, rather than in mammalian neurons. Second, four differential equations must be solved numerically for every unit in the model. As a result, models that include many units could require prohibitive computing time. Third, for most CCN applications the Hodgkin–Huxley model violates the

simplicity heuristic because rarely do such applications attempt to account for data that depend on intracellular concentrations of sodium, potassium, or chloride.

For these reasons, there have been a number of attempts to produce models with fewer equations that display as many of the desirable properties of the Hodgkin–Huxley model as possible. Some of these attempts are described in the following subsections.

6.4.1 The Leaky Integrate-and-Fire Model

The simplest model of a single unit that produces spiking behavior, and also the oldest (Lapique, 1907), is the leaky integrate-and-fire model (e.g., Koch, 1999). Suppose neuron B receives an excitatory projection from neuron A. Let $V_A(t)$ and $V_B(t)$ denote the intracellular voltages at time t in neurons A and B, respectively. Then the leaky integrate-and-fire model assumes that the rate of change of $V_B(t)$ is given by

$$\frac{dV_B(t)}{dt} = \alpha f[V_A(t)] + \beta - \gamma V_B(t), \quad (6.1)$$

where α , β , and γ are constants. The function $f[V_A(t)]$ models temporal delays in the propagation of an action potential from the pre- to the postsynaptic unit. This function is described in detail in Section 6.4.4, but briefly it models temporal delays that occur when an action potential propagates down the axon in the presynaptic unit, and the temporal smearing that occurs during the chemical cascades initiated by this action potential in the synapse and in the dendrites of the postsynaptic unit. The parameter α is a measure of synaptic strength because the larger this value the greater the effect of an action potential in the presynaptic unit. In many applications, learning is modeled by assuming that α changes as a function of experience. The parameter β determines the spontaneous firing rate of unit B, and γ determines the rate at which charged ions leak out of the unit.

Equation (6.1) is a linear differential equation that does not produce spikes. Instead, it predicts continuous and smooth changes in activation. To generate spikes from this model a threshold V_{peak} is set on $V_B(t)$. When $V_B(t)$ exceeds V_{peak} it is reset to V_{reset} and a spike is drawn by hand. An example of activation produced by this model is shown in Figure 6.1. The top panel shows the membrane potential predicted by the model when $V_{peak} = -10$ and $V_{reset} = -50$. The bottom panel adds hand-drawn spikes.

The leaky integrate-and-fire model is simple enough that it can be investigated analytically. In fact, in the absence of any input from the presynaptic neuron A, solving Equation (6.1) shows that $V_B(t)$ is just an exponential function. For almost all other models, however, analytic solutions are unavailable, so numerical solutions are required. In these cases, it can be difficult to predict how the activation will change if any of the numerical constants are changed or if the model is modified in any other way. One solution to this problem is to apply methods from non-linear dynamics (e.g., Strogatz, 2014; Wiggins, 2003).

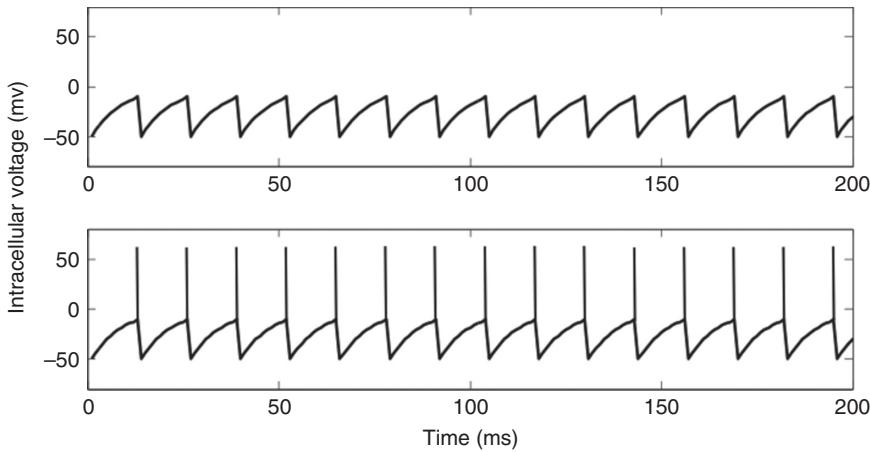


Figure 6.1 Top panel: Activation produced by the leaky integrate-and-fire model (with $\beta = 1/60$, $\gamma = 7/60$, $V_{peak} = -10$, and $V_{reset} = -50$). Bottom panel: The same activation as in the top panel, except with spikes added by hand.

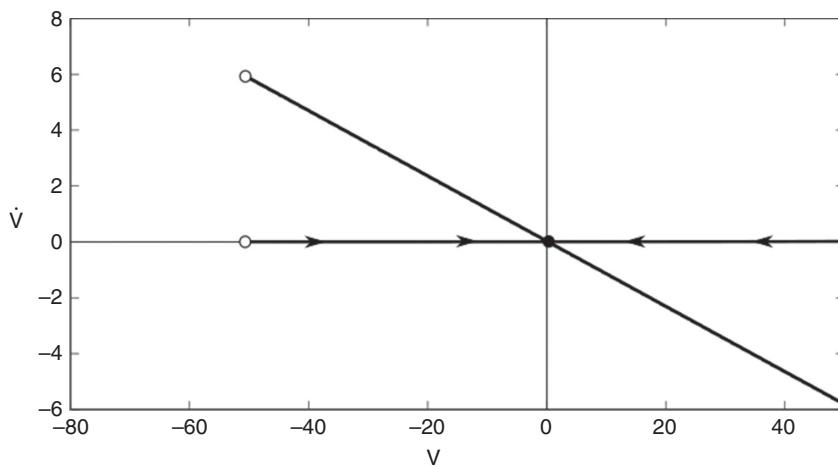


Figure 6.2 Phase portrait for the integrate-and-fire model (with $\beta = 1/60$, $\gamma = 7/60$, $V_{peak} = -10$, and $V_{reset} = -50$). The abscissa is intracellular voltage $V_B(t)$, whereas the ordinate is $\dot{V}_B(t) = \frac{dV_B(t)}{dt}$.

For example, consider Figure 6.2, which shows a sort of phase portrait for the leaky integrate-and-fire model. The abscissa is intracellular voltage, and thus is the same as the ordinate in Figure 6.1. The Figure 6.2 ordinate denotes values of $\frac{dV_B(t)}{dt}$ [denoted there by $\dot{V}_B(t)$]. Thus, voltage increases for any value of V_B for which $\dot{V}_B(t) > 0$ and voltage decreases for any V_B for which $\dot{V}_B(t) < 0$. Note that voltage increases for any negative value of V_B (actually for any $V_B < 1/60$) and decreases when V_B is positive (i.e., when $V_B > 1/60$). When V_B is exactly equal to $1/60$, the

derivative is 0 and, therefore, voltage will remain at this value unless or until some external input is added to the model. Thus, $V_B = 1/60$ is a stable fixed point of the model – or in other words, an attractor.

Figure 6.2 also shows that as voltage increases from the reset value (-50 mv), the derivative continuously decreases, so although voltage increases to the attractor, it does so in a slower and slower manner. In other words, the plot of voltage against time must be negatively accelerating – a property that is easily seen in Figure 6.1.

The fact that the leaky integrate-and-fire model does not naturally predict spiking is widely considered a weakness of the model (e.g., Izhikevich, 2007). Also, it does a relatively poor job of describing ms by ms changes in the membrane potential of real neurons and it is not flexible enough to model qualitative differences in the dynamics of different types of neurons. For these reasons, other single-equation models have been developed.

6.4.2 The Quadratic Integrate-and-Fire Model

Perhaps the most popular single-equation alternative to the leaky integrate-and-fire model replaces the linear decay term with a quadratic polynomial. The resulting model is known as the quadratic integrate-and-fire model (Ermentrout, 1996; Latham *et al.*, 2000). For the scenario modeled in Equation (6.1), the quadratic integrate-and-fire model assumes that the rate of change of $V_B(t)$ is given by

$$\frac{dV_B(t)}{dt} = \alpha f[V_A(t)] + \beta + \gamma [V_B(t) - V_r] [V_B(t) - V_t], \quad (6.2)$$

where α , β , and γ are constants, V_r is the resting membrane potential, V_t is the instantaneous threshold potential and, as before, the function $f[V_A(t)]$ models temporal delays in the propagation of an action potential from one neuron to another. Unlike the leaky integrate-and-fire model, Equation (6.2) produces the upstroke of action potentials via its natural dynamics, although it does not produce the downstroke. To create spikes, an extra voltage resetting step is required to generate the downstroke of the action potential – specifically, when $V_B(t)$ reaches V_{peak} it is reset to V_{reset} . Figure 6.3 shows an example of the spiking behavior produced by Equation (6.2).

The bottom panel of Figure 6.3 shows the model's phase portrait. Note that the derivative $\dot{V}_B(t)$ is always positive, so in this model intracellular voltage can only rise. This is why the artificial voltage resetting mechanism is required. Also note that $\dot{V}_B(t)$ is positively accelerating. Thus, as the voltage increases, the magnitude of the increase becomes progressively greater. This is the property that allows the model to account for the upstroke of the action potential. Because of this property, the quadratic integrate-and-fire model is generally viewed as a superior alternative to the leaky integrate-and-fire model (Izhikevich, 2007).

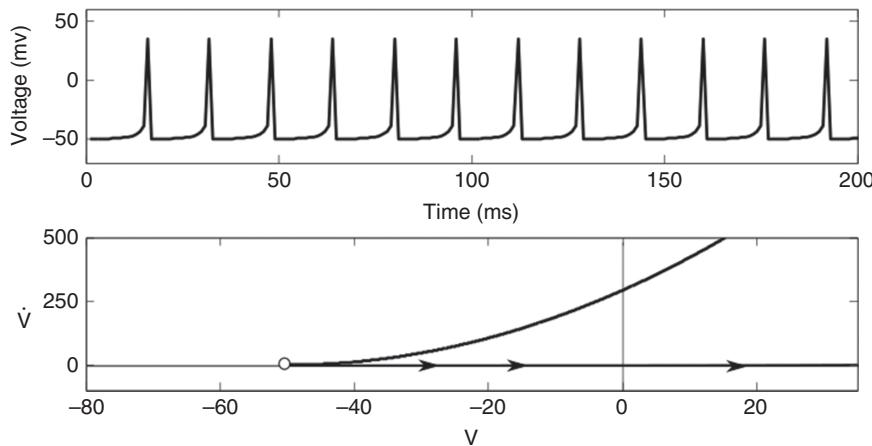


Figure 6.3 Top panel: Typical spiking profile produced by the quadratic integrate-and-fire model. Bottom panel: Phase portrait for the quadratic integrate-and-fire model. In both cases $\beta = 11.83$, $\gamma = .117$, $V_r = -60$, $V_t = -40$, $V_{peak} = 35$, and $V_{reset} = -50$.

6.4.3 The Izhikevich Model

Much more realistic behavior is possible if a second differential equation is added that models slow changes in ion concentrations. One of the first of these was the FitzHugh–Nagumo model, in which the rate of change in voltage (i.e., the derivative) is modeled as a cubic polynomial and slow changes in ion concentrations are modeled with a linear differential equation (FitzHugh, 1961; Nagumo, Arimoto, & Yoshizawa, 1962). Izhikevich (2003) proposed a similar model that replaces the cubic polynomial with the quadratic integrate-and-fire model. The Izhikevich (2003) model requires less computing time to evaluate than the FitzHugh–Nagumo model, has simpler dynamics, and can account for some qualitative firing phenomena that are outside the scope of the FitzHugh–Nagumo model (e.g., tonic and rebound bursting; Izhikevich, 2004). The Izhikevich (2003) model assumes

$$\begin{aligned}\frac{dV_B(t)}{dt} &= \alpha f[V_A(t)] + \beta + \gamma [V_B(t) - V_r] [V_B(t) - V_t] - \theta U_B(t), \\ \frac{dU_B(t)}{dt} &= \lambda [V_B(t) - V_r] - \omega U_B(t),\end{aligned}\quad (6.3)$$

where the quadratic integrate-and-fire model is as before and θ , λ , and ω are constants. In these equations $V_A(t)$ and $V_B(t)$ again denote intracellular voltages at time t and $U_B(t)$ is an abstract regulatory term that is meant to describe slow recovery in unit B after an action potential is generated. $U_B(t)$ could represent activation in the K⁺ current or inactivation in the Na⁺ current, or some combination of both. As before, when $V_B(t)$ reaches V_{peak} it is reset to V_{reset} . At the same time, however, $U_B(t)$ is also reset to $U_B(t) + U_{reset}$.

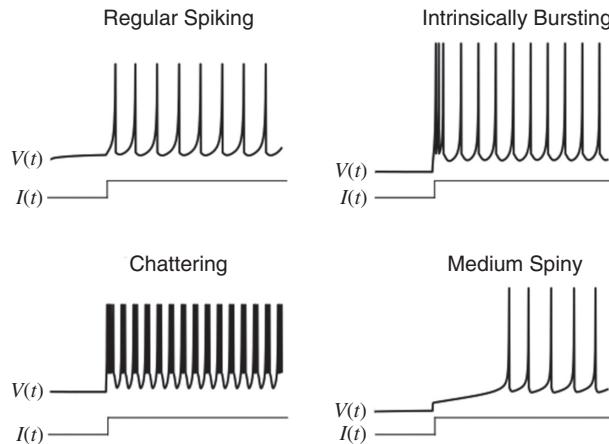


Figure 6.4 Spike trains produced by four different versions of the Izhikevich (2003) model in the same patch-clamp experiment, where the injected current is denoted by $I(t)$. (Regular spiking neuron: $\beta = .52, \gamma = .007, \theta = .01, \lambda = -.06, \omega = 0.03, V_r = -60, V_t = -40, V_{peak} = 35, V_{reset} = -50, U_{reset} = 100$. Intrinsic bursting neuron: $\beta = .52, \gamma = .012, \theta = .01, \lambda = .05, \omega = 0.01, V_r = -75, V_t = -45, V_{peak} = 50, V_{reset} = -56, U_{reset} = 130$. Chattering neuron: $\beta = 1.04, \gamma = .03, \theta = .02, \lambda = .09, \omega = 0.03, V_r = -60, V_t = -40, V_{peak} = 25, V_{reset} = -40, U_{reset} = 150$. Medium spiny neuron: $\beta = 2, \gamma = .02, \theta = .02, \lambda = -.2, \omega = 0.01, V_r = -80, V_t = -25, V_{peak} = 40, V_{reset} = -55, U_{reset} = 150$.)

The Equation (6.3) model is highly flexible and produces some extremely realistic spiking behavior. Figure 6.4 shows examples of four qualitatively different kinds of dynamical behavior that can be produced from this model (from Izhikevich, 2003) when different numerical values are chosen for its parameters, and Izhikevich (2003) has identified at least 17 other types. Figure 6.4 shows predictions from the various models in the same patch-clamp type experiment where current is injected into the unit at the same time in each case. Especially when noise is added, many of the spike trains produced by the models are almost indistinguishable from single unit recordings collected from real neurons (for many examples, see chapter 8, Izhikevich, 2007). One reasonable strategy, which follows from the simplicity heuristic, is to use the Izhikevich model for any units in the network for which single-unit recording data are available. If no such data are available then the simpler quadratic integrate-and-fire model could be used instead. Numerical solutions of Equations (6.1)–(6.3) are readily obtained using Euler's method. For example, Izhikevich (2007) provides Matlab code that solves Equations (6.3) using this approach.

The Izhikevich model has many free parameters that each affect the resulting dynamics of the model – often in interrelated ways. As a result, finding parameter estimates for a neuron type that has not been previously modeled can be a daunting task. Studying the phase portrait can greatly facilitate this process. For

example, consider Figure 6.5, which illustrates the phase portrait of the regular spiking neuron. The top panel shows the spike train produced by the model in a patch-clamp experiment in which current is injected into the neuron beginning at time $t = 900$ ms and ending at time $t = 1200$ ms. Notice that the neuron fires a burst to the injected current, and then immediately transitions to its slow spontaneous firing (with rate completely determined by the parameter β). The middle and bottom panels show phase portraits that describe the model's qualitative behavior. The middle panel shows the phase portrait while the current is being injected and the bottom panel shows the phase portrait after the current has been turned off. In

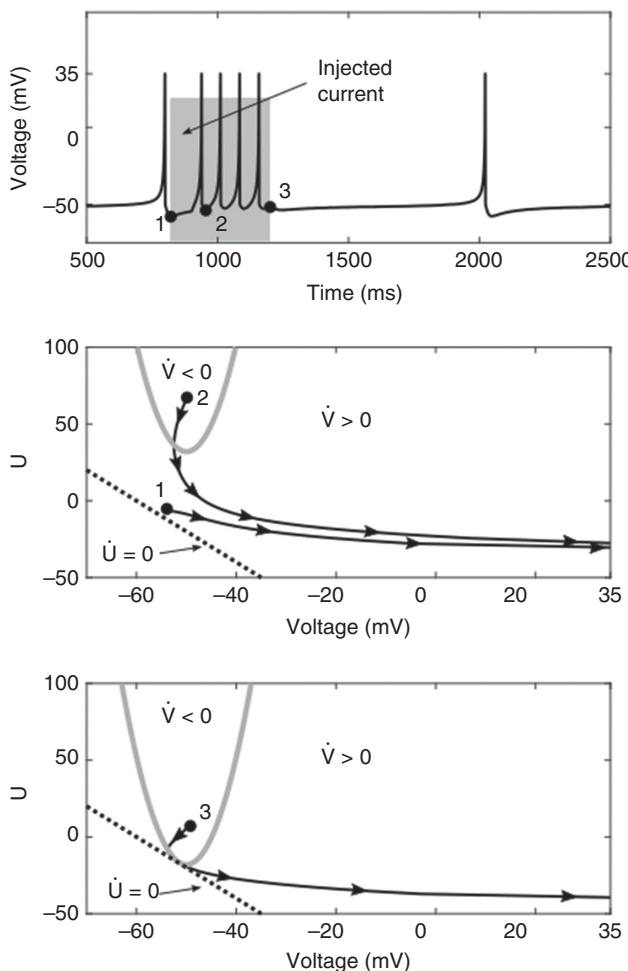


Figure 6.5 Top panel: Typical spiking profile produced by the regular spiking neuron version of the Izhikevich (2003) model in a patch-clamp experiment where current is injected into the unit (denoted by the gray box). Middle panel: Phase portrait during the time of the injected current. Bottom panel: Phase portrait during the time after the injected current.

both figures the abscissa shows values of the intracellular voltage $V_B(t)$ and the ordinate shows values of $U_B(t)$. The gray quadratic curve, known as the voltage nullcline, denotes values of the (V_B, U_B) ordered pair for which $\dot{V}_B(t) = 0$. The derivative is negative for all points inside the parabola and positive for all outside points. The dotted black line is the U_B nullcline [i.e., the set of all (V_B, U_B) for which $\dot{U}_B(t) = 0$], with $\dot{U}_B(t) < 0$ for all points above the line and $\dot{U}_B(t) > 0$ for all points below.

The numbers, 1, 2, and 3, in Figure 6.5 identify time points. Time 1 is at 900 ms – right when the current is first injected. Note that the voltage has just been reset because the model has just spontaneously spiked. Time 2 is just after the voltage was reset following the first spike produced during current injection, and time 3 coincides with the offset of the injected current. Note that at time point 1, $\dot{V}_B(t) > 0$, so the intracellular voltage $V_B(t)$ begins to increase. The further the trajectory moves away from the voltage nullcline, the greater the value of $\dot{V}_B(t)$ and the faster the voltage increase. When the threshold for another spike is reached (i.e., at $V_B(t) = V_{peak}$, where $V_{peak} = 35$ mV in this case), the voltage is reset to $V_B(t) = V_{reset}$ (-50 mV in Figure 6.5) and $U_B(t)$ is reset to its current value plus U_{reset} (i.e., 100). At time 2, note that $\dot{V}_B(t) < 0$, so right after the second spike, voltage decreases slightly (readily apparent in the top panel). However, as the trajectory crosses the voltage nullcline, $\dot{V}_B(t)$ switches to positive and voltage begins to increase again, ever more rapidly until another spike is produced.

Note that the effect of the injected current is to raise the voltage nullcline. This moves more (V_B, U_B) points into the region for which $\dot{V}_B(t) > 0$, which increases the model's firing rate. Thus, manipulating any parameters that affect the width or location of the voltage nullcline will affect the overall firing rate of the model (i.e., γ , V_r , V_t , and β). The delay between spikes can be manipulated via the resetting parameters (V_{reset} and U_{reset}) and the parameters that control $U_B(t)$ (i.e., λ and ω). For example, to increase the delay, parameter values should be chosen that cause the trajectory to reset near the voltage nullcline. This will guarantee that $\dot{V}_B(t)$ is near zero, and therefore that voltage will initially change only slowly.

6.4.4 Modeling Synaptic Delays

Regardless of which spiking-neuron model is used, the free parameters that determine the dynamics of each unit should be set so that the behavior of the unit is as consistent as possible with what is known about the behavior of the real neurons the unit is meant to model. Then, by the set-in-stone ideal, these parameter values should remain invariant across all applications of the model.

The integrate-and-fire model, the quadratic integrate-and-fire model, and the Izhikevich model all describe changes in membrane potential at one particular spatial location within a neuron. They do not describe the propagation of action potentials throughout the cell. Nor do they model delays that occur when an action potential is propagated across a synapse. Modeling these phenomena is considerably more complex.

The standard approach to modeling the propagation of action potentials within a neuron (e.g., down an axon) is called compartment modeling (e.g., Koch, 1999). As mentioned earlier, in this approach a neuron is modeled as a collection of cylinders and spheres, each of which is called a compartment. Separate partial differential equations are written that describe the propagation of the action potential within each compartment and all these equations are used to predict how an action potential propagates from a dendrite down to the end of an axon. The standard partial differential equation that describes propagation within each compartment is called the cable equation (e.g., Koch, 1999). This approach is widely used to account for detailed results of patch-clamp experiments in which current is injected at one location on the neuron and the results are recorded at various other locations.

Modeling synaptic events is potentially even more complex. For example, when an action potential reaches the terminal end of an axon: (1) synaptic vesicles open, (2) neurotransmitter is released, (3) the neurotransmitter diffuses across the synapse, (4) the neurotransmitter binds to postsynaptic receptors, and (5) the neurotransmitter–receptor complex either allows positively charged ions to flow directly into the neuron (in the case of ionotropic receptors) or else initiates a chemical cascade that indirectly causes the postsynaptic neuron to become depolarized (in the case of metabotropic receptors). All these processes cause temporal delay in the propagation of the action potential from the presynaptic to the postsynaptic neuron, and they also introduce a temporal smear. The action potential is a spike, but its postsynaptic effects are not. Modeling any one of these processes can be challenging.¹

Models that describe how action potentials propagate down an axon and cross synapses must necessarily be considerably more complex than the one- and two-equation spiking neuron models considered in this chapter. The benefit of this extra complexity is that such models can make predictions at a more reductionistic level than, say, the Izhikevich model. The Izhikevich model gives good accounts of spike trains, but is unable to account for data at lower levels (e.g., patch-clamp experiments that record from more than one site on the neuron; experiments that measure changes in ion concentrations over time). The cost of the extra complexity needed to make predictions at lower levels is a dramatic increase in computing time. One goal of CCN applications is to account for at least some behavior. Thus, CCN models must include multiple units in multiple brain regions. As a result, computing time is a serious consideration. The current state-of-the-art in CCN modeling is to account for any data between behavior (at the highest level) and spike trains (at the lowest level). In such applications, compartment models and models of synaptic transmission would be used only to predict the time-course of the postsynaptic effects of a spike in the presynaptic neuron. If this is the goal then we should seek simpler alternatives.

¹ Note that most applications of compartment modeling do not model any of these synaptic events. This is because the most common goal of compartment modeling is to account for results of single-neuron patch-clamp experiments.

The problem is to model the temporal delays of spike propagation and the temporal smearing that occurs at the synapse in a simple way that can be combined with any of the one- or two-equation models of spiking considered above. A standard solution is to use the so-called alpha function (Rall, 1967). This is the function $f[V_A(t)]$ in Equations (6.1)–(6.3). The idea is that every time the presynaptic unit spikes, the following input is delivered to the postsynaptic unit (with spiking time $t = 0$):

$$f(t) = \frac{t}{\delta} \exp\left(\frac{\delta - t}{\delta}\right). \quad (6.4)$$

This function has a maximum value of 1.0 and it decays to .01 at $t = 7.64\delta$. Thus, δ can be chosen to model any desired temporal delay. If a second spike occurs before $f(t)$ decays to zero then a second alpha function is added to the residual $f(t)$ (again, with time of the second spike at $t = 0$).

6.4.5 Noise

In many applications it will be desirable to add noise to the models. There are two primary advantages to this. First, of course, human behavior is almost always probabilistic. Without noise the models will always produce the same result given the same initial conditions. So noise is necessary to account for probabilistic responding. Second, noise can drive the model out of a dangerous attractor state that can arise with any global learning rule.

Learning rules in neural network models can be classified as local or global. Local rules, like backpropagation, modify every synapse using a different error signal. In contrast, global learning rules use the same error signal at every synapse. The evidence is good that most of the brain uses global learning rules (with the possible exception of the cerebellum). For example, dopamine (DA), which is widely thought to serve as a training signal, is released in roughly equal amounts at all target synapses. One dangerous property of global learning is that it can lead to an attractor state in which response accuracy is constrained to remain at chance. Fortunately, adding noise to a model can break it free from this dangerous attractor state.

As an illustration of this problem, consider a simple two-stimulus, two-response task in which the network must learn to emit one response if stimulus A is presented and another response if stimulus B is presented. So after training, presentation of stimulus A should activate motor unit A more strongly than motor unit B. Initially, before training, the strength of the synapse between sensory unit A and the two motor units should be roughly equal. If two synaptic strengths are exactly equal, then the pre- and postsynaptic activations will be identical at both synapses (since the presynaptic activation is from the same cortical unit), and therefore any global learning algorithm will specify an equal amount of strengthening or weakening of both synapses on every trial, regardless of whether the response was correct or incorrect. Thus, if there is no noise, then once the weights become equal

they must remain equal for all time, thereby preventing the network from learning the desired associations. Adding noise to the postsynaptic activation breaks the model free from this attractor state. When noise is added, the postsynaptic activations at the two synapses will not be the same, even if the presynaptic activations and synaptic strengths are identical. As long as the postsynaptic activations are different, the change in synaptic strength will be different at the two synapses and learning can proceed.

Noise can be added to each of the models by adding a white noise term to each voltage equation. For example, the leaky integrate-and-fire model then becomes

$$\frac{dV_B(t)}{dt} = \alpha f [V_A(t)] + \beta - \gamma V_B(t) + \sigma \epsilon(t), \quad (6.5)$$

where σ is a constant and $\epsilon(t)$ is white noise. Note that except for the first, input term, Equation (6.5) is exactly equivalent to an Ornstein–Uhlenbeck process. So whereas $\epsilon(t)$ and $\epsilon(t + \tau)$ are statistically independent for any value of τ , $V_B(t)$ and $V_B(t + \tau)$ are not independent, at least for reasonably small values of τ (because of the leak created by the $\gamma V_B(t)$ term). In real biological systems, inertia prevents physical changes large enough to guarantee that $\epsilon(t)$ and $\epsilon(t + \tau)$ will be independent for small values of τ . As a result, many researchers have proposed that the Ornstein–Uhlenbeck process is a better model of biological noise processes than the Wiener process (e.g., Ricciardi & Sacerdote, 1979).

Adding a white noise term to each voltage equation converts the differential equation into a stochastic differential equation (e.g., Øksendal, 2003). In general, this complicates the process of deriving numerical predictions from the model. First, of course, adding noise to a voltage equation introduces variability to the spike times, which will generally cause variability in the predicted value of every dependent variable. This variability complicates model evaluation (for details see Section 6.8).

Second, the white noise term will defeat many numerical algorithms that are commonly used to solve differential equations. This is because white noise is not smooth and many differential equation solvers dynamically adjust the step size depending on the smoothness of the solution. In the presence of white noise, these algorithms keep shrinking the step size in an unsuccessful attempt to find a smooth solution, and often will eventually fail because the solution is not smooth even when the step size is as small as possible. On the other hand, simpler methods that use a fixed step size – such as Euler’s method – work well in the presence of noise.

6.5 Firing-rate Models

The human brain contains somewhere on the order of 10^{11} neurons. Although Izhekovich actually constructed a spiking-neuron model with this many units,² simulating any behavioral task with this complex a model is essentially

² This unpublished simulation, which made no attempt to simulate behavior, is described on his website at www.izhikevich.org/human_brain_simulation/Blue_Brain.htm.

impossible with today's technology. Furthermore, the brain is thought to exhibit considerable redundancy. For example, many neurons in the same cortical column or hypercolumn exhibit similar firing properties. For reasons such as these, an alternative approach to spiking-neuron models tries to model mean activity in large populations of neurons. The key variable is the instantaneous firing rate within each of these populations, and as a result, this class of models is known as firing-rate models.

Firing-rate models originated with the pioneering work of Wilson and Cowan (1972, 1973), who based their work on mean-field approaches from statistical mechanics. A complete review of this large field is beyond the scope of this chapter. For more details, see the excellent chapters by Ermentrout and Terman (2010, chapter 11) or Dayan and Abbott (2001, chapter 7).

Firing-rate models make many strong assumptions. In particular, they assume that all neurons within a population (e.g., a cortical column) are statistically identical and in close spatial proximity of each other. In addition, it is assumed that the neurons within a population are randomly interconnected and that these interconnections are dense enough so that any two neurons in the same population are connected – either directly or via interneurons. Furthermore, each population is assumed to include many neurons, and only two kinds of populations are allowed – one in which all neurons are excitatory and another in which all neurons are inhibitory. Given these assumptions, the mean instantaneous firing rate in any single population is defined as

$$R(t) = \lim_{\Delta T \rightarrow 0} \frac{1}{\Delta T} \left(\frac{\# \text{ of spikes during } (t, t + \Delta T) \text{ in population}}{\# \text{ of neurons in population}} \right). \quad (6.6)$$

Note that as $\Delta T \rightarrow 0$, each neuron in the population either fires once during the interval $(t, t + \Delta T)$ or not at all. Therefore, $R(t)$ can also be interpreted as the proportion of active neurons at time t , and as a result $0 \leq R(t) \leq 1$.

As mentioned earlier, a popular method for modeling the postsynaptic effects of a presynaptic spike is via the alpha function described by Equation (6.4). According to this model, every presynaptic spike generates a new alpha function postsynaptically. Firing-rate models follow a similar approach. Suppose the firing rate in a presynaptic population of excitatory neurons is $R_{\text{pre}}(t)$, and further suppose that the mean synaptic strength between all neurons in the pre- and postsynaptic populations is w . Then in firing-rate models, postsynaptic activation at time t , denoted by $I_{\text{post}}(t)$, is equal to the convolution of the presynaptic firing rate $R_{\text{pre}}(t)$ and the alpha function, weighted by mean synaptic strength:

$$I_{\text{post}}(t) = w \int_0^t f(t-s) R_{\text{pre}}(s) ds, \quad (6.7)$$

where $f(t)$ is the alpha function of Equation (6.4).

This approach is easily extended to more complex architectures. For example, suppose some postsynaptic population receives input from M_E populations of excitatory neurons and M_I populations of inhibitory neurons. Let w_{Ei} denote the mean

strength of all synapses of excitatory presynaptic population i neurons onto neurons in the postsynaptic population, and let w_{Ij} denote the mean strength of all synapses of inhibitory presynaptic population j neurons onto postsynaptic neurons. Then firing-rate models predict that the mean activation in the postsynaptic population equals:

$$I_{\text{post}}(t) = \sum_{i=1}^{M_E} w_{Ei} \int_0^t f(t-s) R_{Ei}(s) ds - \sum_{j=1}^{M_I} w_{Ij} \int_0^t f(t-s) R_{Ij}(s) ds, \quad (6.8)$$

where $R_{Ei}(t)$ is the firing rate in excitatory presynaptic population i at time t and $R_{Ij}(t)$ is the firing rate in inhibitory presynaptic population j .

The Equation (6.4) alpha function models postsynaptic effects of a spike [i.e., $f(t)$ in Equations (6.7) and (6.8)] as a gamma function. In firing-rate models, other choices are also used. An especially popular choice – largely because of mathematical tractability – is the exponential function $f(t) = (1/\tau)e^{-t/\tau}$. In this case, Equation (6.8) reduces to

$$\tau \frac{dI_{\text{post}}(t)}{dt} = \sum_{i=1}^{M_E} w_{Ei} R_{Ei}(t) - \sum_{j=1}^{M_I} w_{Ij} R_{Ij}(t) - I_{\text{post}}(t) \quad (6.9)$$

Note the similarity of this equation to the leaky integrate-and-fire model described in Equation (6.1). The main difference is that artificial spikes are generated in the leaky integrate-and-fire model but not in Equation (6.9).

The last step is to convert the postsynaptic activation into postsynaptic firing rate. This is necessary because the activations predicted by Equations (6.8) and (6.9) are theoretically unbounded, whereas the firing rate $R(t)$ is constrained to the interval $[0,1]$. The standard approach is to assume that the postsynaptic firing rate equals

$$R_{\text{post}}(t) = F[I_{\text{post}}(t)], \quad (6.10)$$

where F is a monotonically increasing function known as the activation function. A common choice is to assume that F is a (sigmoidal) logistic function. According to this model,

$$R_{\text{post}}(t) = \frac{1}{1 + \exp\left(-\frac{I_{\text{post}}(t)-\alpha}{\beta}\right)}, \quad (6.11)$$

where α and β are constants. The constants α and β can be used to model a non-zero tonic firing rate and that presynaptic firing does not cause postsynaptic firing rate to saturate. When $\alpha = 0$, note that $R_{\text{post}}(t) = .5$ if the activation is 0 (i.e., if $I_{\text{post}}(t) = 0$). Thus, when $\alpha = 0$ the tonic firing rate is substantial (i.e., half the neurons in the population are active at any given time). Larger values of α cause the tonic firing rate to decrease. In contrast, increasing the value of β lowers the asymptotic firing rate produced by any given (constant) activation level.

As an example, consider the simple case where a single presynaptic population of excitatory neurons projects to a postsynaptic population. Figure 6.6 shows

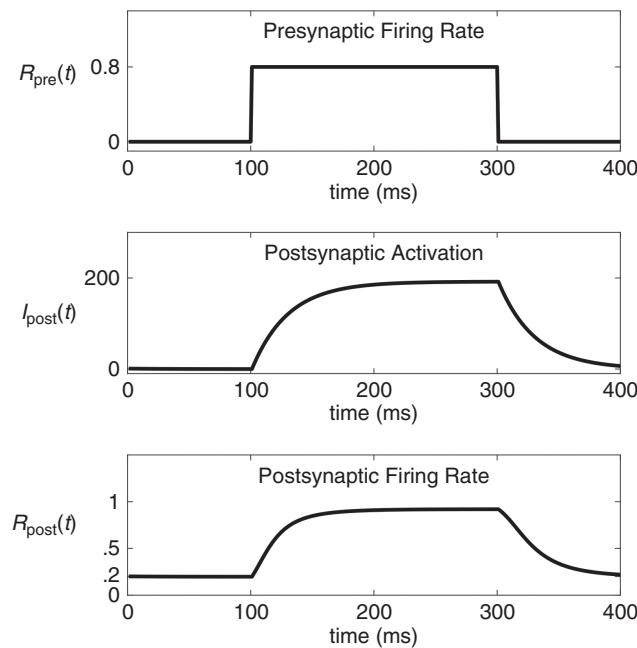


Figure 6.6 Predictions of a firing-rate model for the case where an excitatory presynaptic population projects to a postsynaptic population. The top panel shows the firing rate of the presynaptic population. The middle panel shows the postsynaptic activation, and the bottom panel shows the postsynaptic firing rate (with $\tau = 30$, $w_{Ej} = 240$, $\alpha = 70$, $\beta = 50$).

predictions of Equations (6.9) and (6.11) when the firing rate of the presynaptic population equals .8 during the time interval [100,300] and is zero at all other times. Several features of these predictions deserve comment. First, note that the spontaneous firing rate equals 0 in the presynaptic population and .2 in the postsynaptic population (because $\alpha > 0$) – an illustration that firing-rate models can account for any tonic firing rate. Second, note that the model makes the reasonable prediction that a sudden onset of firing in the presynaptic population causes the postsynaptic firing rate to increase gradually (i.e., exponentially), and similarly that a sudden drop in presynaptic firing causes a gradual decrease (again exponential) in postsynaptic firing rate.

6.6 Learning

6.6.1 Synaptic Plasticity

The many different learning algorithms that can be found in the machine learning literature are often classified into three types. *Unsupervised learning* algorithms operate in the absence of any feedback or guidance. The most widely known

version is Hebbian learning, in which all active synapses are strengthened, regardless of whether the response was rewarded or punished. *Reinforcement learning* algorithms depend on feedback but not on any guided instruction. They typically use a global learning algorithm that delivers the same feedback signal to every synapse in the network. *Supervised learning* algorithms depend on local feedback that supplies a unique training signal to every synapse. The most widely known examples include the delta rule and backpropagation.

The human brain exhibits many forms of neural plasticity that operate over a wide range of different timescales (for reviews, see, e.g., Malenka & Siegelbaum, 2001; Stanton, Bramham, & Scharfman, 2006). The plasticity-related phenomena that are widely thought to be associated with learning and memory are long-term potentiation (LTP) and long-term depression (LTD) (e.g., Martin, Grimwood, & Morris, 2000). LTP and LTD refer to a long-lasting increase and decrease, respectively, in the efficacy of a synapse, which results from simultaneously stimulating the pre- and postsynaptic neurons. LTP and LTD have been closely studied in many different brain regions and in many different cell types. The conditions that promote LTP and LTD are qualitatively different in different regions, and it has been noted that these conditions closely match popular unsupervised-, reinforcement-, and supervised-learning algorithms. In particular, Doya (2000) and others have noted that the rules that govern LTP and LTD match Hebbian learning in the cortex and medial temporal lobe structures (including hippocampus; e.g., Feldman, 2009), reinforcement learning in the basal ganglia (and especially the striatum; e.g., Houk, Adams, & Barto, 1995), and supervised learning in the cerebellum.

The most common excitatory neurotransmitter in the brain is glutamate and virtually all long-range cortical projections are glutamatergic. For these reasons, it is fortunate that the most widely studied form of LTP occurs at glutamatergic synapses. Glutamate binds to a number of different types of receptors, but the most important for LTP are NMDA receptors. NMDA is an ionotropic receptor that is a channel for Na^+ and Ca^{2+} . It requires partial depolarization to become activated (because of an extracellular Mg^{2+} plug that prevents Na^+ and Ca^{2+} from entering the cell during resting membrane potentials), and so it has a higher threshold for activation than other non-NMDA glutamate receptors (e.g., AMPA).

NMDA-receptor activation initiates a number of chemical cascades that can affect synaptic plasticity. One of the most important and best understood is the pathway that phosphorylates calcium/calmodulin-dependent protein kinase II (CaMKII). When calcium enters the cell through the activated NMDA receptor, it binds to calmodulin and the calcium/calmodulin complex phosphorylates CaMKII. When fully phosphorylated, CaMKII initiates a variety of processes that eventually increase the efficacy of the synapse (e.g., Lisman, Schulman, & Cline, 2002). Dopamine plays an important role in these processes because if it is in the synapse within a second or so of the NMDA-receptor activation then it can potentiate the phosphorylating effects of calcium/calmodulin (via D1 receptor activation) and thereby potentiate synaptic efficacy (Yagishita *et al.*, 2014).

A large literature shows that DA neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNpc) increase their firing above baseline following unexpected rewards (e.g., Hollerman & Schultz, 1998; Mirenowicz & Schultz, 1994; Schultz, 1998). Thus, this form of DA-enhanced LTP should be in effect following an unexpected reward in any brain region that is a target of VTA or SNpc DA neurons (and that expresses DA D1 receptors). This includes all of the frontal cortex but not, for example, the visual or auditory cortex. In these regions, however, there is evidence that acetylcholine may play a modulatory role similar to DA in LTP and LTD (e.g., Gu, 2003; McCoy, Huang, & Philpot, 2009). A variety of evidence suggests that the long-term efficacy of the synapse is weakened (i.e., LTD occurs) when presynaptic activation either fails to activate NMDA receptors, or else activates them only weakly (Bear & Linden, 2001; Kemp & Bashir, 2001).

Although the biochemistry of CaMKII-mediated synaptic plasticity is similar in all DA target regions, the functional role of this plasticity is qualitatively different in the striatum and frontal cortex. Within the striatum, DA is quickly cleared from synapses by DA active transporter (DAT) and, as a result, the temporal resolution of DA in the striatum is high enough for DA to serve as an effective reinforcement-learning signal. For example, if the first response in a training session is correct and the second response is an error then within the striatum, the elevated DA levels that result from the positive feedback on trial 1 should have decayed back to baseline levels by the time of the response on trial 2. Unlike the striatum, however, DAT concentrations in the frontal cortex are low (e.g., Seamans & Robbins, 2010). As a result, cortical DA levels change slowly. For example, the delivery of a single food pellet to a hungry rat increases DA levels in the prefrontal cortex (PFC) above baseline for approximately 30 min (Feenstra & Botterblom, 1996). Thus, the first rewarded behavior in a training session is likely to cause frontal cortical DA levels to rise, and the absence of DAT will cause DA levels in the frontal cortex to remain high throughout the training session. As a result, all synapses that are activated during the session are likely to be strengthened, regardless of whether the associated behavior is appropriate. Thus, although DA may facilitate LTP in the frontal cortex, it appears to operate too slowly to serve as a frontal–cortical reinforcement training signal (Lapish *et al.*, 2007).

6.6.2 Models of Learning in the Striatum and Cortex

The structural changes that accompany LTP and LTD can be modeled in a variety of ways. One critical decision is whether to build a discrete-time or a continuous-time model. This choice largely depends on the nature of the data that the model will be tested against. If the data have a discrete trial-by-trial structure (i.e., the time is reset at the beginning of each trial), as is common in many cognitive-behavioral experiments, then a discrete-time model should be used because no data would exist to test the extra assumptions required of a continuous-time model. On the other hand, when modeling a continuous-time task (i.e., when the time is reset only once, typically at the beginning of the experiment), a continuous-time

learning model is required. A cognitive example might be a sequence-learning task in which feedback is provided following each response and there is no pause between responses.

6.6.2.1 Discrete-time Models of Learning at Synapses that Lack Fast DA Reuptake

At synapses that lack fast DA reuptake, synaptic plasticity mimics Hebbian learning. In the frontal cortex, for example, the first rewarded response should cause DA levels to rise above baseline and subsequent rewarded responses will cause DA to remain elevated for the duration of the training session. As a result, all active synapses will be strengthened, regardless of whether they received correct or error feedback. In this case, the key phenomena to model are that plasticity depends only on the product of pre- and postsynaptic activation. Strengthening of the synapse requires post-synaptic NMDA receptor activation. Activation below this threshold weakens the synapse.

Let $w_{A,B}(n)$ denote the strength of the synapse on trial n between presynaptic unit A and postsynaptic unit B, and let $V_J(t)$ denote the intracellular activation in unit J ($J = A$ or B) at time t . The key variables to compute are the integrated alpha functions of units A and B. Suppose the time between stimulus presentation and response is T . Then define

$$I_J(T) = \int_0^T f[V_J(t)]dt \quad (6.12)$$

for $J = A$ or B . Note that $I_J(T)$ describes the total postsynaptic effect of all spikes produced by unit J during the duration of the trial. Given these definitions, the following difference equation can be used to adjust the strength of the A/B synapse between trials n and $n + 1$:

$$\begin{aligned} w_{A,B}(n+1) &= w_{A,B}(n) \\ &+ \alpha H[I_B(t) - \theta_{NMDA}] I_A(t) \left\{ 1 - e^{-\lambda[I_B(t) - \theta_{NMDA}]} \right\} [1 - w_{A,B}(n)] \\ &- \beta H[\theta_{NMDA} - I_B(t)] I_A(t) e^{-\lambda[\theta_{NMDA} - I_B(t)]} w_{A,B}(n). \end{aligned} \quad (6.13)$$

The function $H[g(x)]$ is the Heaviside function that equals 1 when $g(x) \geq 0$ and 0 when $g(x) \leq 0$. The constant θ_{NMDA} represents the threshold for NMDA-receptor activation. When postsynaptic activation is right at this threshold then the unit will produce a certain number of spikes during the trial duration T . Each spike generates an alpha function and θ_{NMDA} is theoretically equal to the integral of all these superimposed alpha functions. So the synaptic strengthening term is positive only on trials when the postsynaptic activation consistently exceeds the threshold for NMDA-receptor activation. When the synapse is strengthened, note that the amount of strengthening increases with the product of the presynaptic activation and an exponentially increasing function of the postsynaptic activation – similar to all other versions of Hebbian learning. The $[1 - w_{A,B}(n)]$ term is a rate-limiting

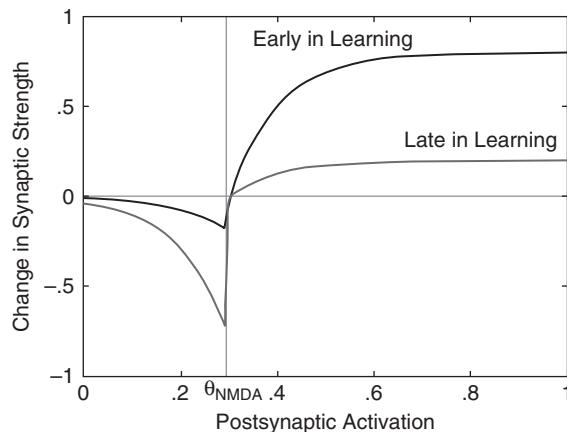


Figure 6.7 Change in synaptic strength predicted by the Hebbian learning model described in Equation (6.13) as a function of amount of postsynaptic activation (here scaled from 0 to 1). Predictions are shown for early in learning [i.e., when $w_{A,B}(n) = 0.2$] and late in learning [i.e., when $w_{A,B}(n) = 0.8$].

term that prevents $w_{A,B}(n + 1)$ from exceeding 1, and the constant λ scales the postsynaptic activation.

Most Hebbian learning rules do not include any mechanism to decrease synaptic strength. The last term in Equation (6.13) is therefore unusual. First, note that this last term equals 0 except when postsynaptic activation is consistently below the NMDA-receptor threshold. Second, note that the exponential term reaches its maximum when postsynaptic activation is near the NMDA threshold and decreases as the postsynaptic activation gets smaller and smaller. This is consistent with the neurobiology. For example, in the absence of any postsynaptic activation we do not expect any synaptic plasticity. The $w_{A,B}(n)$ at the end prevents $w_{A,B}(n + 1)$ from dropping below 0.

Figure 6.7 shows predicted changes in synaptic strength [i.e., $w_{A,B}(n + 1) - w_{A,B}(n)$] for this model as a function of the magnitude of postsynaptic activation during both early [when $w_{A,B}(n) = 0.2$] and late [when $w_{A,B}(n) = 0.8$] learning.

6.6.2.2 Discrete-time Models of Learning at Synapses with Fast DA Reuptake

In the striatum, DA reuptake is fast, so at cortical–striatal synapses LTP and LTD follow a form of reinforcement learning. One way to model synaptic plasticity at such synapses is as follows:

$$\begin{aligned} w_{A,B}(n + 1) &= w_{A,B}(n) + \alpha H[I_B(t) - \theta_{\text{NMDA}}] H[D(n) - D_{\text{base}}] \\ &\quad \times I_A(t) \left\{ 1 - e^{-\lambda[I_B(t) - \theta_{\text{NMDA}}]} \right\} [D(n) - D_{\text{base}}][1 - w_{A,B}(n)] \\ &\quad - \beta H[I_B(t) - \theta_{\text{NMDA}}] H[D_{\text{base}} - D(n)] \end{aligned}$$

$$\begin{aligned} & \times I_A(t) \left\{ 1 - e^{-\lambda[I_B(t) - \theta_{NMDA}]} \right\} [D_{\text{base}} - D(n)] w_{A,B}(n) \\ & - \gamma H [\theta_{NMDA} - I_B(t)] I_A(t) e^{-[\theta_{NMDA} - I_B(t)]} w_{A,B}(n), \end{aligned} \quad (6.14)$$

where $D(n)$ is the amount of DA released on trial n and D_{base} is the baseline DA level.

Note that the synaptic strengthening term requires two conditions – postsynaptic activation above the threshold for NMDA-receptor activation and DA above baseline. Once these conditions are met, synaptic strengthening is the same as in the Equation (6.13) Hebbian learning model. Two different conditions cause the synapse to be weakened. The second [the last γ term in Equation (6.14)] is the same as in the Hebbian learning model. The first (i.e., the β term), however, is unique to striatal-mediated reinforcement learning. Cortical–striatal synapses are weakened if postsynaptic activation is strong and DA is below baseline – a condition that would occur, for example, on trials when feedback indicates the trial n response was incorrect.

Figure 6.8 shows predicted changes in synaptic strength [i.e., $w_{A,B}(n+1) - w_{A,B}(n)$] for this model as a function of the magnitude of postsynaptic activation, separately for early [when $w_{A,B}(n) = 0.2$] and late [when $w_{A,B}(n) = 0.8$] learning,

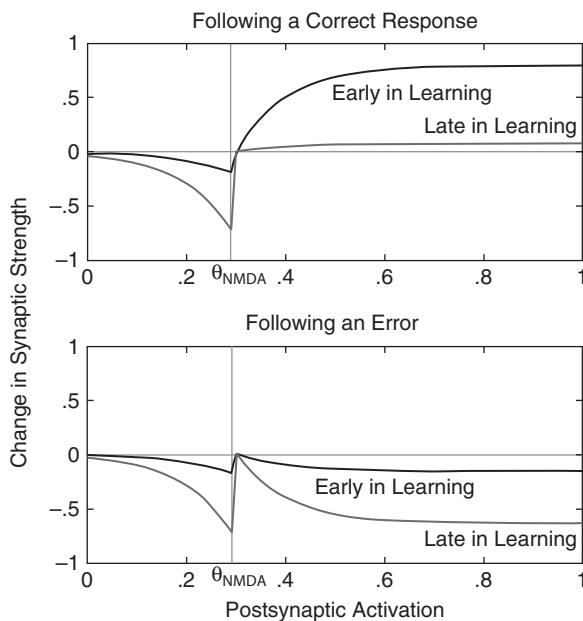


Figure 6.8 Change in synaptic strength predicted by the reinforcement learning model described in Equation (6.14) as a function of amount of postsynaptic activation (here scaled from 0 to 1). Predictions are shown for early in learning [i.e., when $w_{A,B}(n) = 0.2$] and late in learning [i.e., when $w_{A,B}(n) = 0.8$], and following a correct response and an error ($\alpha = 2$, $\beta = 4$, $\gamma = 1$).

and following correct and incorrect responses. Note that synaptic plasticity following correct (rewarded) responses is similar to plasticity in the Hebbian model (compare the top panel of Figure 6.8 with Figure 6.7). The only real difference is that learning is attenuated more during late learning in the reinforcement learning model. This is because DA fluctuations decrease as rewards become more predictable (more on this in the next subsection). Note that errors have a greater effect on synaptic plasticity late in learning. This is because errors are expected early in learning, so DA fluctuations are small. Late in learning, however, when accuracy is high, errors are unexpected, which causes a large DA depression and therefore a large decrease in synaptic efficacy.

6.6.2.3 Modeling DA Release

The Equation (6.14) model of reinforcement learning requires that we specify the amount of DA released on every trial in response to the feedback signal [the $D(n)$ term]. The more that DA increases above baseline (D_{base}), the greater the increase in synaptic strength, and the more it falls below baseline, the greater the decrease.

Although there are a number of powerful models of DA release, Equation (6.14) requires only that we specify the amount of DA released to the feedback signal on each trial. The key empirical results are (e.g., Schultz, Dayan, & Montague, 1997; Tobler, Dickinson, & Schultz, 2003): (1) midbrain DA neurons fire tonically, and therefore have a non-zero baseline (i.e., spontaneous firing rate); (2) DA release increases above baseline following unexpected reward, and the more unexpected the reward the greater the release; and (3) DA release decreases below baseline following unexpected absence of reward, and the more unexpected the absence, the greater the decrease. One common interpretation of these results is that over a wide range, DA firing is proportional to the reward prediction error (RPE) – that is, to the difference between obtained reward and predicted reward. If we denote the obtained reward on trial n by R_n and the predicted reward by P_n , then the RPE on trial n is defined as:

$$RPE_n = R_n - P_n. \quad (6.15)$$

So positive prediction errors occur when the reward is better than expected, and negative prediction errors when the reward is worse than expected. Either signals that learning is incomplete.

A simple model of DA release can be built by specifying how to compute (1) obtained reward, (2) predicted reward, and (3) exactly how the amount of DA release is related to the RPE. A straightforward solution to these three problems is as follows (Ashby & Crossley, 2011). First, in tasks that provide positive feedback, negative feedback, or no feedback on every trial and where reward magnitude never varies, then a simple model can be used to compute obtained reward. Specifically, define the obtained reward R_n on trial n as +1 if correct or reward feedback is received, 0 in the absence of feedback, and -1 if error feedback is received.

Second, predicted reward can be computed using a simple average of past rewards, so long as the average employs temporal discounting in order to ensure

that recent trials are weighted more heavily than earlier trials. Temporal discounting is critical to make the model sensitive to abrupt changes in reward probabilities. According to this approach, the predicted reward on trial $n + 1$ equals

$$P_{n+1} = \frac{1}{S_n} \sum_{i=1}^n \theta^{n-i} R_i, \quad (6.16)$$

where

$$S_n = \sum_{i=1}^n \theta^{i-1}.$$

So S_n is the sum of the weights on each R_i . Note that if $\theta = 1$ then $S_n = n$ and P_n equals the arithmetic mean of all past rewards. The more θ is reduced below 1, the greater the temporal discounting. Note that this model of predicted reward is stimulus- and response-dependent. Because predicted reward can vary greatly across stimuli and typically depends on what response is emitted, “trial $n + 1$ ” in Equation (6.16) should be interpreted as the $(n + 1)^{\text{th}}$ occurrence of this same stimulus and response.

Equation (6.16) is not the most convenient form for computing predicted reward – in part because it requires starting the sum from scratch on every trial. A more convenient form would be one where the current estimate of predicted reward is updated after each new reward is received. It turns out that Equation (6.16) can be rewritten in such a manner as follows:

$$P_{n+1} = P_n + \frac{1}{S_n} (R_n - P_n). \quad (6.17)$$

Equations in this form are ubiquitous in the reinforcement learning literature (e.g., Sutton & Barto, 1998) – that is, equations where the new estimate (e.g., P_{n+1}) is constructed by adjusting the old estimate (e.g., P_n) by a fractional amount of the prediction error (i.e., $R_n - P_n$). Models based on this form include the Rescorla–Wagner Model, temporal difference learning, SARSA, and Q-learning (e.g., see Sutton & Barto, 1998). Because of its great importance, a derivation of Equation (6.17) is given next.

Derivation of Equation (6.17) Equation (6.17) is derived from Equation (6.16) as follows:

$$\begin{aligned} P_{n+1} &= \frac{1}{S_n} \sum_{i=1}^n \theta^{n-i} R_i \\ &= \frac{1}{S_n} \left(R_n + \sum_{i=1}^{n-1} \theta^{n-i} R_i \right) \\ &= \frac{1}{S_n} \left(R_n + \theta \sum_{i=1}^{n-1} \theta^{n-1-i} R_i \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{S_n} \left(R_n + \frac{\theta S_{n-1}}{S_{n-1}} \sum_{i=1}^{n-1} \theta^{n-1-i} R_i \right) \\
&= \frac{1}{S_n} (R_n + \theta S_{n-1} P_n) \\
&= \frac{1}{S_n} R_n + \frac{\theta S_{n-1}}{S_n} P_n.
\end{aligned}$$

Next, note that $\theta S_{n-1} = S_n - 1$. Therefore,

$$\begin{aligned}
P_{n+1} &= \frac{1}{S_n} R_n + \frac{S_n - 1}{S_n} P_n \\
&= \frac{1}{S_n} R_n + P_n - \frac{1}{S_n} P_n,
\end{aligned}$$

from which Equation (6.17) follows immediately. \square

The final problem is to determine the amount of DA release associated with every possible value of RPE_n . A simple model was proposed by Ashby and Crossley (2011), who assumed that the amount of DA release is related to the RPE in a manner that is consistent with the data reported by Bayer and Glimcher (2005). Specifically, they assumed that

$$D(n) = \begin{cases} 1 & \text{if } RPE_n > 1 \\ .8RPE_n + .2 & \text{if } -.25 \leq RPE_n \leq 1 \\ 0 & \text{if } RPE_n < -.25. \end{cases} \quad (6.18)$$

Note that the baseline DA level is .2 (i.e., when $RPE_n = 0$) and that DA levels increase linearly with the RPE. However, note also the asymmetry between DA increases and decreases. As is evident in the Bayer and Glimcher (2005) data, a negative RPE quickly causes DA levels to fall to zero, whereas there is a considerable range for DA levels to increase in response to positive RPEs.³

6.6.2.4 Continuous-time Models of Hebbian Learning

All the learning models considered so far assume the data come from an experiment with a discrete-trial structure. In this case, synaptic strengths are updated off-line between trials. However, in continuous-time tasks the updating must be done in real time. This requires more detail than in the models we have so far considered. Not surprisingly, fewer continuous-time models of learning have been proposed.

³ Bayer, Lau, and Glimcher (2007) subsequently reported that when the RPE is negative, DA firing remains below baseline for longer periods the more negative the RPE, suggesting that negative RPEs may be coded by a combination of firing rate and the duration of the pause in DA cell firing. This suggests that the dynamic range of positive and negative RPEs may be more balanced than assumed by the Equation (6.18) model.

Even so, one continuous-time model of Hebbian learning is widely used. This model was motivated by evidence that the magnitude and even the direction of plasticity at a synapse depends not only on the magnitude of the pre- and postsynaptic activations, but also on the timing – a phenomenon known as spike-timing-dependent plasticity (STDP). Considerable data show that if the postsynaptic neuron fires just after the presynaptic neuron then synaptic strengthening (i.e., LTP) occurs, whereas if the postsynaptic neuron fires first then the synapse is weakened (e.g., Bi & Poo, 2001; Sjöström *et al.*, 2008). Furthermore, the magnitude of both effects seems to fall off exponentially as the delay between the spikes in the pre- and postsynaptic neurons increases. Let T_{pre} and T_{post} denote the time at which the pre- and postsynaptic neurons fire, respectively. Then a popular model of STDP (e.g., Zhang *et al.*, 1998) assumes that the amount of change in the synaptic strength equals

$$\Delta = \begin{cases} e^{-\theta_+(T_{\text{post}} - T_{\text{pre}})}, & \text{if } T_{\text{post}} > T_{\text{pre}} \\ -e^{\theta_-(T_{\text{post}} - T_{\text{pre}})}, & \text{if } T_{\text{post}} < T_{\text{pre}}. \end{cases} \quad (6.19)$$

Figure 6.9 shows an example of this function.

To implement this form of Hebbian learning, the strength of each synapse is updated according to Equation (6.19) anytime the pre- and postsynaptic units both fire.

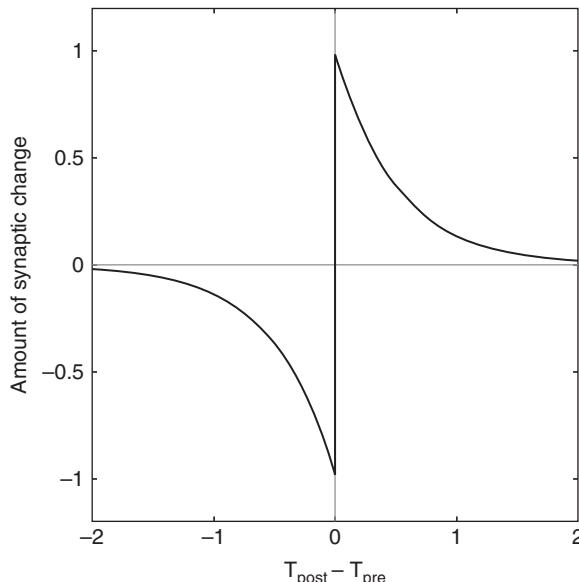


Figure 6.9 Amount of change in synaptic strength predicted by STDP as a function of the difference in time between firing in the postsynaptic neuron (i.e., T_{post}) and the presynaptic neuron (i.e., T_{pre}).

6.7 Testing CCN Models

CCN models can be tested against a wide variety of data, including data from single-unit recording, fMRI, and TMS experiments, as well as behavioral experiments with either healthy young adults or certain special neuropsychological patient groups (e.g., Parkinson's disease patients), who may or may not be operating under the influence of some drug (e.g., a DA agonist or antagonist). In most of these cases, some modeling interface is required to generate the relevant dependent measure from the neural activations that the models produce in each of their included brain regions. This section describes the most common of those interfaces and discusses some special issues that might arise during the modeling process.

6.7.1 Single-unit Recording Data

One advantage that the spiking-neuron models hold over the firing-rate models is that only the former can be tested against single-unit firing data. A two-step process is recommended. During the first step, the models of each neuron type are separately fitted to spike trains collected from single-neuron patch-clamp experiments. For example, if the Izhikevich spiking model is used then the end result of this first step will be numerical values for all constants in Equation (6.3) that allow the model to provide good fits to the patch-clamp data. An example is shown in Figure 6.10.

Because of all the free parameters that are estimated, this step does not provide a test of the model. For example, because the patch-clamp data are from a single neuron, the network architecture of the model is irrelevant to goodness-of-fit. Even so, this step achieves two important goals. First, if successful, it guarantees that the units of the model have similar qualitative dynamics to the neurons they represent. Second, it fixes most of the free parameters of the model. By the set-in-stone ideal, all free parameters that are fixed during this step must remain at these same fixed values during all future tests of the model. This is a major advantage of spiking-unit models over firing-rate models. In most cases, all free parameters of firing-rate models must be estimated during the model-testing process, whereas most free parameters of spiking-unit models can be estimated during this first, preliminary step. Therefore, when fit to behavioral or fMRI data, spiking-unit models will typically require estimation of fewer free parameters than firing-rate models.

During the second step, the model is tested against single-unit recordings from the same neuron types as in the first step, except during *in vivo* recordings made while the animal is engaged in a behavior as similar as possible to the behavior that is the main focus of study. Since these recordings depend on the network architecture, this step provides a test of the model. The test is not parameter-free, however, because some parameters will remain unestimated after step 1. Mostly these will be synaptic strengths between different connected units in the model.

For example, Ashby and Crossley (2011) used this approach to test a CCN model of striatal function. They first used patch-clamp data to build accurate Izhikevich models of two prominent striatal neuron types – namely, medium spiny neurons

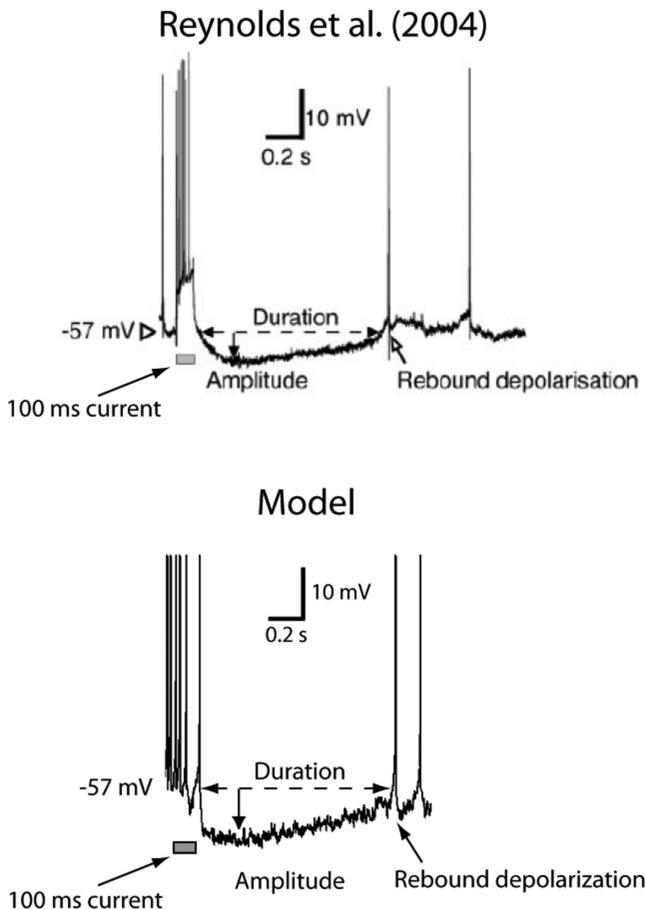


Figure 6.10 Patch-clamp recording data from the striatal TAN of a rat (top panel) and fits of a modified Izhikevich model under the same experimental conditions (from Ashby & Crossley, 2011).

(96% of all striatal neurons) and TANs (i.e., tonically active neurons, which represent 2% of all striatal neurons; the TAN fit is shown in Figure 6.10). Next, with all these parameters held fixed, they showed that their network model accurately accounted for *in vivo* recordings from medium spiny neurons and TANs in a variety of different behavioral paradigms.

6.7.2 Behavioral Data

Spiking-unit and firing-rate networks produce activation in a distributed neural network, but without some added assumptions they produce no behavior. So to fit the models to behavioral data, it is necessary to add some assumptions that describe how neural activation is related to behavior. In most cases, this process involves three steps. The first is to identify which brain region in the hypothesized network

controls the behavioral response – that is, one must decide where to place the decision units. The second step is to decide, in each unit, what function of neural activity should drive the decision. For example, should the decision be based on the number of spikes, or the spiking rate, or perhaps on the integrated membrane potential? Finally, in tasks with multiple response alternatives, the third step is to decide how to resolve the competition among the various competing units in the critical brain region.

Step 1. What brain region controls behavior? The decision about where to place the decision units depends on one's knowledge of the task and the relevant neuroscience literature, and on one's modeling goals. In tasks that require finger or arm movements, typical choices would be the supplementary motor area, dorsal or ventral premotor cortex, or primary motor cortex. In contrast, if the task requires an eye-movement response then the critical area may be in the lateral intraparietal area, the supplementary eye fields, the frontal eye fields, or the superior colliculus. On the other hand, in many cases the goal may be to model cognition rather than the specific motor response that implements the outcome of the relevant cognitive processes. Ignoring motor processing simplifies the modeling because all areas downstream of the critical cognitive region can be omitted. Note that this strategy will underestimate RT since some key synapses will be omitted, but it might not affect accuracy predictions at all, especially in tasks where errors are due to cognitive failures, rather than to simple motor errors. For example, models of working memory typically assume that the key decision units are in the PFC (e.g., Ashby *et al.*, 2005; Frank, Loughry, & O'Reilly, 2001), since an extensive literature implicates the PFC as the most critical site for working memory. As a result, models of working memory often grossly oversimplify or omit altogether projections from PFC to premotor and motor cortices.

Step 2. What function of neural activity drives the decision? After the anatomical location of the decision units has been selected, the next step is to decide what function of activity in these units will initiate the behavior. With firing-rate models the obvious choice is to set a threshold on firing rate. When the threshold is crossed the behavior is initiated. With spiking-unit models, several choices are possible, but one especially appealing choice is to set a threshold on the integrated output alpha function:

$$\int_0^t f[V_B(x)] dx, \quad (6.20)$$

where $f[\cdot]$ is the alpha function defined in Equation (6.4). The idea is to compute this integral continuously and initiate the behavior when the threshold is first exceeded. This decision variable has a number of attractive properties. Most importantly, it depends on the unit's output and because we expect the motor response to be driven by the output of the units in the decision region, the integrated alpha function is therefore as close to the behavior as possible without adding another downstream unit to the model (e.g., in contrast to a choice such as intracellular voltage, or even the number of spikes produced).

Step 3. How is a response selected when there are multiple alternatives?

There has been considerable work on this problem in the field of neuroscience over the past decade or so. Especially illuminating have been studies in which single-unit recordings were made from putative decision neurons during a task in which an animal had to select among competing motor responses on each trial (for reviews, see, e.g., Bogacz *et al.*, 2010; Rangel & Hare, 2010; Wang, 2008). For example, in an early and influential study, Shadlen and Newsome (2001) reported that neurons in the lateral intraparietal area reliably predicted the eye-movement responses of monkeys in a task that required the animals to determine the direction of motion of random dot patterns. Furthermore, these neurons displayed the push–pull profile that one might expect from a classic diffusion process – that is, neurons that predicted a movement of the eyes to the right increased their firing rate when the correct response to the stimulus was a rightward movement and decreased their firing rate when the stimulus signaled a leftward movement. The formal correspondence between these properties and the diffusion process was quickly noted (e.g., Smith & Ratcliff, 2004).

Of course, generalizing the diffusion model to more than two alternatives is not straightforward, but it is well known that an accumulator or race model with lateral inhibition among the channels mimics a diffusion process (Bogacz *et al.*, 2007; Usher & McClelland, 2001). Thus, in tasks with more than two response alternatives, a sound yet reasonably simple solution is to set a criterion on each decision unit and allow the first unit that crosses this threshold to control the response, but also to build in lateral inhibition among all decision units (McMillen & Holmes, 2006; Usher & McClelland, 2001).

For example, suppose each of M decision units is modeled via the quadratic integrate-and-fire model described by Equation (6.2). Suppose the J th of these units receives input from m units in an earlier layer (or brain structure). Then we can model the lateral inhibition among output units on unit J via:

$$\frac{dV_J(t)}{dt} = \alpha \sum_i^m f[V_i(t)] + \beta + \gamma [V_B(t) - V_r] [V_B(t) - V_t] - \omega \sum_{I \neq J}^M f[V_I(t)]. \quad (6.21)$$

The last term is a standard model of lateral inhibition (e.g., Usher & McClelland, 2001). Note that this model assumes that the total amount of lateral inhibition on unit J is an increasing function of the total amount of activation in all output units.

6.7.3 fMRI Data

One of the great advantages of CCN models over traditional cognitive models is that they can be tested against fMRI and other neuroscience data. CCN models predict changes in neural activation in a variety of different brain regions and fMRI records an indirect measure of neural activation. Thus, fMRI provides a natural platform from which to test CCN models. Nevertheless, some challenges

must be overcome to take full advantage of fMRI methodology. First, CCN models make direct predictions about neural activation, but they do not make direct predictions about the blood-oxygen-level-dependent (BOLD) signal that is most commonly measured in fMRI experiments. Thus, the first problem is to generate predicted BOLD responses from the model's predicted neural activations. Second, CCN models make anatomic predictions about where the task-related changes in neural activation should be found, but they typically do not make predictions that are specific enough to identify a small set of voxels in the region of interest (ROI) that could be used to test the model. For example, a model might specify that during a certain period of a working memory task, a specific type of activation should occur in dorsolateral PFC (dlPFC). However, such a model would generally not predict that every voxel in dlPFC would show this activation pattern – only that some would. So, a second significant problem that must be solved is to identify exactly which voxels within dlPFC should be used to test the model. Finally, a third problem is to compare the observed and predicted BOLD responses in the selected voxels and to decide on the basis of this comparison whether the model succeeds or fails at accounting for the results of the experiment.

Although a variety of similar solutions to these problems have been proposed (Ashby & Waldschmidt, 2008), within the past few years consensus has settled on an approach, called model-based fMRI, that fully exploits current fMRI data-analysis software packages (O'Doherty, Hampton, & Kim, 2007). The basic idea is to first fit the model to the behavioral data collected during the scanning session separately for each participant. Next, parameter estimates from the model fits are used to generate predicted neural activations that are unique for every participant. The third step is to generate a predicted BOLD response from each brain region in the model, and then to correlate these predictions with observed BOLD responses from every voxel in the brain. These two steps can be performed by any of the popular fMRI data-analysis software packages. Finally, all the resulting correlations are assessed for statistical significance.

Model-based fMRI can be used to account for individual differences in fMRI data, but if the computational model is good, it can also be used to identify brain regions that respond selectively to components or sub-processes of the task. In particular, if the model has different parameters that describe different perceptual or cognitive processes that are presumed to mediate the behavior under study, then different regressors can be created that make specific predictions about each of these processes. For example, O'Doherty *et al.* (2004) used this approach to identify separate brain regions associated with the actor versus the critic in actor-critic models of reinforcement learning.

Steps 1 and 2. Generating predicted neural activations. The first step is to fit the model separately to the behavioral data collected from each participant during the scanning session. Details on this process are given in Section 6.8. After this fitting process is complete, each participant will be characterized by a unique set

of parameter estimates. The second step is to use these estimates to generate predicted neural activations for each brain region identified by the model. The critical issue to consider here is spatial resolution. A typical voxel size in fMRI is 2^3 to 3^3 mm 3 . So the goal of this step should be to produce the model's best estimate of total neural activation in regions of about this size. This typically means that activations from all neurons within the same specified brain region should be added together.

Logothetis and colleagues reported evidence that the BOLD response is more closely related to local field potentials than to the spiking output of individual neurons (Logothetis, 2003; Logothetis *et al.*, 2001). Local field potentials integrate the field potentials produced by small populations of cells over a sub-millimeter range, and they vary continuously over time. So if the CCN model is constructed from spiking units, the spike trains produced by the model must be converted to local field potentials. This can be done by low-pass filtering each spike. Fortunately, this is exactly the operation performed by the alpha function. So with spiking-unit models, after the scanning task is simulated (separately for each participant) every spike within each brain region is used to trigger an alpha function and these are all added together to mimic the spatial summation that occurs during fMRI. These summed alpha functions represent the predicted local field potentials during the scanning session within the corresponding brain region. Predicted activation in firing-rate models is already temporally smoothed, so with firing-rate models no extra low-pass filtering is needed.

Step 3. Generating predicted BOLD responses. The fMRI BOLD response increases with the amount of oxygenated hemoglobin in a voxel relative to the amount of deoxygenated hemoglobin (Ogawa *et al.*, 1990). In comparison with the neural activation that presumably drives it, the BOLD response is highly sluggish, reaching a peak around 6 s after the neural activation that induced it, and slowly decaying back to baseline 20–25 s later. Almost all current applications of fMRI assume that the transformation from neural activation to BOLD response can be modeled as a linear, time-invariant system (e.g., Boynton *et al.*, 1996).

In the linear systems approach, one can conceive of the vascular system that responds to a sudden oxygen debt as a black box in which the input is neural activation, and the output is the BOLD response. If the system is linear and time-invariant, then it is well known that the BOLD response at time t , denoted by $B(t)$, to any neural activation $N(t)$ can be written as

$$B(t) = \int_0^t N(x)h(t-x)dx. \quad (6.22)$$

Equation (6.22) is the well-known convolution integral that completely characterizes the behavior of any linear, time-invariant system (see, e.g., Chen, 1970). The function $h(t)$ is traditionally called the impulse response function because it describes the response of the system to an input that is a perfect impulse. In the fMRI literature, however, $h(t)$ is known as the hemodynamic response function,

often abbreviated as hrf. The hrf is the hypothetical BOLD response to an idealized impulse of neural activation, typically peaking at 6 s and lasting for 30 s or so. Thus, predicted BOLD responses can be generated from the model by numerically convolving the predicted neural activations with a suitable model of the hrf. Popular fMRI data analysis software packages such as SPM and FSL will perform the numerical convolution. All the user needs to do is supply a vector that contains the predicted neural activation at each TR of the scanning session and specify a functional form for the hrf. Many alternative models of the hrf have been proposed (e.g., a common choice is a gamma function), and the packages allow the user considerable flexibility with respect to this choice. For a complete description of this entire process, see Ashby (2011).

Step 4. Comparing predicted and observed BOLD responses. Once predicted BOLD responses have been computed, the next step is to correlate these predictions with the observed BOLD responses in every voxel. All popular fMRI software packages routinely compute these correlations by using the general linear model (GLM) of statistics (see Ashby, 2011, for details). The standard analysis converts each correlation coefficient to the z - or t -statistic associated with the null hypothesis that the correlation is zero. The result is a z -statistic (for example) in every voxel in the brain, which collectively are known as a statistical parametric map (SPM).

Step 5. Assessing statistical significance. The final step is to make a statistical significance decision about every z -statistic in the SPM. Of course, with only one such decision the solution to this problem is taught in every introductory statistics course, but spatial resolution is high enough with modern imaging equipment that an adult human brain might be characterized by several hundred thousand voxels, and therefore several hundred thousand simultaneous significance decisions are required. To complicate matters further, spatial correlations guarantee that the separate statistics are not independent. Although there is no optimal solution to this problem, many different alternative methods have been proposed for correcting for this huge number of multiple comparisons. Current software packages allow the user to choose among many of these alternative solutions (again, see Ashby, 2011, for details).

This correlational analysis can be used in either a confirmatory or an exploratory manner. The confirmatory analysis is to check whether significant correlations appear in the brain regions predicted by the model. Most CCN models will predict different neural activations in each brain region included in the model. The strongest possible confirmatory result would be that voxels in which the observed BOLD response is significantly correlated with the predicted BOLD response in one of these regions appear in that region but not in any of the other regions in the model, and that a similar unique confirmation is found for every hypothesized brain region. So for a model that includes brain regions A, B, and C, voxels where the BOLD response correlates with the predicted BOLD response in region A are found in region A, but not in regions B or C, and voxels where the BOLD response

correlates with the predicted BOLD response in region B are found in region B, but not in regions A or C (and similarly for region C).

The exploratory analysis is to identify brain regions that are not in the model in which the observed BOLD response is nevertheless correlated with the predicted BOLD response for some region in the model. Such correlations could exist for a number of reasons, including because the model is incomplete.

6.7.4 TMS Data

TMS uses a small device to direct focused electromagnetic pulses through the skull of human participants. The stimulation typically targets a specific cortical site that was previously identified using high-resolution MRI. Many studies have used TMS to investigate the causal role of some specific cortical region in a particular cognitive process or behavior (e.g., Sandrini, Umiltà, & Rusconi, 2011).

Most current TMS studies use theta-burst stimulation, with typical protocols delivering a burst of 3 pulses at 50 Hz (i.e., 20 ms between each pair of pulses) and then repeating this pattern for as long as several minutes. Theta-burst TMS induces electrophysiological changes in the targeted site that last up to 60 min, and behavioral changes in tasks that depend on that site for a similar time period (Huang *et al.*, 2005). The current thinking is that theta-burst TMS induces long-term changes in synaptic plasticity within targeted regions by altering the pattern of Ca^{2+} influx through postsynaptic NMDA receptors (Huang *et al.*, 2011). Whether synaptic strengthening or weakening are potentiated depends on the exact stimulation protocol. For example, a continuous 40-s train of 3-pulse bursts has a long-term inhibitory effect, whereas a 2-s train of 3-pulse bursts that is repeated every 10 s for 3 min causes a long-term facilitatory effect. Huang *et al.* (2011) developed a mathematical model that accurately predicts the long-term effects of a wide variety of different theta-burst protocols.

In most TMS experiments the data of primary interest are behavioral (i.e., RTs and accuracies). The goal is typically to investigate how the TMS affects these behavioral measures. CCN models that assign a functional role to the cortical region targeted by the TMS can be tested against the resulting data. A particularly simple way to model the effects of theta-burst TMS within the CCN framework that is nevertheless consistent with the Huang *et al.* (2011) theory is to assume that TMS changes θ_{NMDA} in Equations (6.13) and (6.14) – that is, it changes the threshold for NMDA-receptor activation. Increasing θ_{NMDA} simulates a reduction in Ca^{2+} influx and will cause more synaptic weakening and less strengthening. In contrast, decreasing θ_{NMDA} simulates an increase in Ca^{2+} influx and causes less synaptic weakening and more strengthening. For example, Hélie *et al.* (2015) used this approach to show that their proposed CCN model of automatic sequence production successfully accounted for the interfering effects of TMS to the supplementary motor area on the RT speed-up that normally occurs during sequence

learning (e.g., where the TMS data were reported by Verwey, Lammens, & Honk, 2002).

6.7.5 Pharmacological and Neuropsychological Patient Data

Many CCN models also make predictions about how performance should change in relevant tasks when participants perform the task under the influence of certain medications or drugs, or when the participants are from some special neuropsychological population. For example, because the reinforcement learning model described in Equation (6.14) includes a term that depends on the amount of DA released on each trial, any CCN model that uses this reinforcement learning algorithm should make specific predictions about how performance should be affected by any drug or neuropsychological condition that alters brain DA levels. This would include DA agonists and antagonists, and neuropsychological conditions such as Parkinson's disease. For example, Hélie, Paul, and Ashby (2012a, 2012b) used this approach to account for a variety of cognitive deficits that occur during Parkinson's disease and to account for the beneficial effects of positive mood on rule-based category learning.

6.8 Parameter Estimation and Model Evaluation

It is almost always impossible to derive predictions from CCN models analytically. Even the simplest CCN models are typically described by a rather large set of simultaneous non-linear differential equations. At best these can be solved numerically. However, most CCN models will include noise terms, in which case the differential equations become stochastic. In this large majority of cases, Monte Carlo simulation is almost always necessary.

Finding best-fitting values of the parameters is a notoriously difficult problem when the model predictions are noisy and require simulation. The typical approach is to simulate many independent and identical subjects and then treat the mean of all these simulations as the model predictions. This can be time-consuming and of course the resulting predictions will still be noisy and therefore standard minimization algorithms are not appropriate. Fortunately, genetic algorithms (e.g., Haupt & Haupt, 2004) can often be used successfully. For example, Cantwell, Crossley, and Ashby (2015) estimated parameters of a spiking-unit CCN model using particle swarm optimization (Clerc, 2012), which creates a population of potential solutions (the “particles”) and then iteratively moves these particles in parameter space according to both their historically best position, and the best-known position of their neighborhood. Due to the stochastic nature of the models, the “function” to be optimized (e.g., sum of squared errors) is not strictly a function at all. Hence, particle swarm optimization, which makes very few assumptions about the form of the problem, is an appropriate tool where traditional optimization routines will fail. After parameter estimation was complete, Cantwell *et al.* (2015) ran an additional 100 simulations with the best-fitting parameter values

and the model predictions were computed by taking the mean across all 100 simulations.

The inflexibility of CCN models (i.e., see Section 6.2) eases the parameter estimation process. Small changes in almost any parameter usually cause only a negligible change in the model's predictions, and therefore small errors in the parameter estimation process will generally have little or no effect on any conclusions that are drawn about the empirical validity of the model. For example, following a crude parameter estimation process, Ashby and Crossley (2011) implemented a sensitivity analysis in which the most important parameters in their spiking-unit CCN model were successively changed by -1% , -10% , $+1\%$, and $+10\%$. After each change, the behavior of the model was simulated under the same experimental conditions that were used to generate the data the model was tested against. Next, after each new simulation, the correlation was computed between the predictions of the best-fitting model and the predictions generated from the new version of the model. In all except one case, these correlations exceeded .99, suggesting that the model makes the same qualitative predictions for a wide range of each of its parameters. The only exception occurred for a $+10\%$ increase in a response threshold parameter [i.e., the threshold mentioned in the discussion of Equation (6.20)]. In this case, the correlation was .74. Importantly, however, even in this case, the perturbed model predicted the same qualitative pattern to the data as the best-fitting version of the model.

This inflexibility also means that parameter space partitioning (PSP; Pitt *et al.*, 2006) will often allow stronger inferences when applied to CCN models than when applied to traditional process-level models. PSP determines what different kinds of qualitative data patterns a model can predict by systematically exploring the model's entire parameter space. Specifically, it uses an efficient Markov chain Monte Carlo search algorithm to compute the volume of parameter space over which the model can account for each different qualitative data pattern.

PSP is an especially effective method for rejecting a model because if observed data show a particular qualitative pattern and the PSP indicates that the volume of parameter space where the model can mimic this pattern is 0, then the model can be rejected. Of course, flexible models can fit more different types of data patterns, so the more flexible the model the less that can be learned from PSP. With CCN models, however, PSP can lead to some strong inferences. For example, Valentin, Maddox, and Ashby (2016) used PSP to show that CCN models based on current theories of DA release are incompatible with observed effects of aggregate feedback on procedural learning. Similarly, Paul and Ashby (2013) used PSP to show that a large class of CCN models could only account for empirical interactions between declarative and procedural memory systems by predicting the existence of an as-yet-undiscovered neuroanatomical projection from the cortex to the striatum.

6.9 Conclusions

The birth of mathematical psychology did not signal the advent of mathematical modeling in psychology. For example, Fechner, Thurstone, and Hull all incorporated heavy doses of modeling into their research programs. Even so, these were all descriptive models, or in the language of Marr (1982), computational-level models. The birth of mathematical psychology could be seen as the beginning of process- or algorithmic-level modeling in psychology. Stimulus sampling theory (Estes, 1950) sparked a revolution – not because it was the first mathematical model of learning (e.g., see Hull, 1943) – but at least in part because it was the first algorithmic-level model. Now, more than a half-century later, it may be time to move to the next level of modeling – namely, the implementational level. CCN represents a serious attempt to take this next step.

CCN presents new challenges to mathematical psychology. Some basic knowledge of neuroscience is required and parameter estimation tends to be more difficult than with many traditional process models. Even so, the potential benefits are significant. These include (1) model convergence, because different researchers must respect similar neuroscience constraints, (2) faster rejection of poor models, because of the mathematical inflexibility of CCN models, (3) the ability to test models against a much wider spectrum of data types, and (4) the potential to unite disparate fields when common brain regions are implicated in seemingly unrelated behaviors.

Acknowledgment

Preparation of this chapter was supported by National Institutes of Health Grant 2R01MH063760.

References

- Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Sciences*, 12(4), 136–143.
- Ashby, F. G. (2011). *Statistical analysis of fMRI data*. Cambridge, MA: MIT Press.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105(3), 442–481.
- Ashby, F. G., & Crossley, M. J. (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience*, 23(6), 1549–1566.
- Ashby, F. G., Ell, S. W., Valentin, V. V., & Casale, M. B. (2005). FROST: A distributed neurocomputational model of working memory maintenance. *Journal of Cognitive Neuroscience*, 17(11), 1728–1743.
- Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, 31(7), 1114–1125.

- Ashby, F. G., & Hélie, S. (2011). A tutorial on computational cognitive neuroscience: Modeling the neurodynamics of cognition. *Journal of Mathematical Psychology*, 55(4), 273–289.
- Ashby, F. G., & Valentín, V. V. (2007). Computational cognitive neuroscience: Building and testing biologically plausible computational models of neuroscience, neuroimaging, and behavioral data. In M. J. Wenger & C. Schuster (eds.), *Statistical and process models for cognitive neuroscience and aging* (pp. 15–58). Mahwah, NJ: Erlbaum.
- Ashby, F. G., & Waldschmidt, J. G. (2008). Fitting computational models to fMRI data. *Behavior Research Methods*, 40(3), 713–721.
- Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1), 129–141.
- Bayer, H. M., Lau, B., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, 98(3), 1428–1439.
- Bear, M., & Linden, D. (2001). The mechanisms and meaning of long-term synaptic depression in the mammalian brain. In W. Cowan, T. Sudhof, & C. Stevens (eds.), *Synapses* (pp. 455–517). Baltimore, MD: Johns Hopkins University Press.
- Bhalla, U. S., & Bower, J. M. (1993). Exploring parameter space in detailed single neuron models: Simulations of the mitral and granule cells of the olfactory bulb. *Journal of Neurophysiology*, 69(6), 1948–1965.
- Bi, G.-q., & Poo, M.-m. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, 24(1), 139–166.
- Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: Multi-alternatives, nonlinearity and value-based multi-dimensional choice. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1485), 1655–1670.
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10–16.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, 16(13), 4207–4221.
- Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin & Review*, 22(6), 1598–1613.
- Cantwell, G., Riesenhuber, M., Roeder, J. L., & Ashby, F. G. (2017). Perceptual category learning and visual processing: An exercise in computational cognitive neuroscience. *Neural Networks*, 89, 31–38.
- Chen, C.-T. (1970). *Introduction to linear system theory*. New York, NY: Holt, Rinehart and Winston.
- Clerc, M. (2012). Standard particle swarm optimisation. *Open access archive HAL*.
- Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to pre-frontal cortex, cognitive control and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 351(1346), 1515–1527.
- Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99(1), 45–77.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

- Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Current Opinion in Neurobiology*, 10(6), 732–739.
- Doya, K. (2007). Reinforcement learning: Computational theory and biological mechanisms. *HFSP Journal*, 1, 30–40.
- Dyson, F. (2004). A meeting with Enrico Fermi. *Nature*, 427(6972), 297–297.
- Ermentrout, G. B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Computation*, 8(5), 979–1001.
- Ermentrout, G. B., & Terman, D. H. (2010). *Mathematical foundations of neuroscience*. New York, NY: Springer Science & Business Media.
- Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57(2), 94–107.
- Feenstra, M. G., & Botterblom, M. H. (1996). Rapid sampling of extracellular dopamine in the rat prefrontal cortex during food consumption, handling and exposure to novelty. *Brain Research*, 742(1), 17–24.
- Feldman, D. E. (2009). Synaptic mechanisms for plasticity in neocortex. *Annual Review of Neuroscience*, 32, 33–55.
- FitzHugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6), 445–466.
- Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *Journal of Cognitive Neuroscience*, 17(1), 51–72.
- Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, 1(2), 137–160.
- Grafton, S. T., Hazeltine, E., & Ivry, R. B. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, 7(4), 497–510.
- Gu, Q. (2003). Contribution of acetylcholine to visual cortex plasticity. *Neurobiology of Learning and Memory*, 80(3), 291–301.
- Hartley, T., Burgess, N., Lever, C., Cacucci, F., & O'Keefe, J. (2000). Modeling place fields in terms of the cortical inputs to the hippocampus. *Hippocampus*, 10(4), 369–379.
- Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms*. New York, NY: John Wiley & Sons.
- Haykin, S. (2009). *Neural networks and learning machines*, third edition. Upper Saddle River, NJ: Prentice Hall.
- Hélie, S., Paul, E. J., & Ashby, F. G. (2012a). A neurocomputational account of cognitive deficits in Parkinson's disease. *Neuropsychologia*, 50(9), 2290–2302.
- Hélie, S., Paul, E. J., & Ashby, F. G. (2012b). Simulating the effects of dopamine imbalance on cognition: From positive affect to Parkinson's disease. *Neural Networks*, 32, 74–85.
- Hélie, S., Roeder, J. L., Vucovich, L., Rünger, D., & Ashby, F. G. (2015). A neurocomputational model of automatic sequence production. *Journal of Cognitive Neuroscience*, 27(7), 1456–1469.
- Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4), 304–309.
- Houk, J., Adams, J., & Barto, A. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. L. Houk J. C. Davis &

- D. G. Beiser (eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT Press.
- Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, 45(2), 201–206.
- Huang, Y.-Z., Rothwell, J. C., Chen, R.-S., Lu, C.-S., & Chuang, W.-L. (2011). The theoretical model of theta burst form of repetitive transcranial magnetic stimulation. *Clinical Neurophysiology*, 122(5), 1011–1018.
- Hull, C. (1943). *Principles of behavior*. New York, NY: Appleton-Century-Crofts.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6), 1569–1572.
- Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5), 1063–1070.
- Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. Cambridge, MA: MIT Press.
- Kemp, N., & Bashir, Z. I. (2001). Long-term depression: A cascade of induction and expression mechanisms. *Progress in Neurobiology*, 65(4), 339–365.
- Koch, C. (1999). *Biophysics of computation: Information processing in single neurons*. New York, NY: Oxford University Press.
- Lapique, L. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization. *Journal of Physiology & Pathology: General*, 9, 620–635.
- Lapish, C. C., Kroener, S., Durstewitz, D., Lavin, A., & Seamans, J. K. (2007). The ability of the mesocortical dopamine system to operate in distinct temporal modes. *Psychopharmacology*, 191(3), 609–625.
- Latham, P. E., Richmond, B., Nelson, P., & Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. Theory. *Journal of Neurophysiology*, 83(2), 808–827.
- Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Reviews Neuroscience*, 3(3), 175–190.
- Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*, 23(10), 3963–3971.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150–157.
- Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review*, 11(5), 945–952.
- Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V., & Ashby, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research*, 74(2), 219–236.
- Malenka, R. C., & Siegelbaum, S. A. (2001). Synaptic plasticity. In W. M. Cowan, T. C. Sudhof, & C. F. Stevens (eds.), *Synapses* (pp. 393–453). Baltimore, MD: Johns Hopkins University Press.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Freeman.
- Martin, S., Grimwood, P., & Morris, R. (2000). Synaptic plasticity and memory: An evaluation of the hypothesis. *Annual Review of Neuroscience*, 23(1), 649–711.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from

- the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McCoy, P. A., Huang, H.-S., & Philpot, B. D. (2009). Advances in understanding visual cortex plasticity. *Current Opinion in Neurobiology*, 19(3), 298–304.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133.
- McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1), 30–57.
- Meeter, M., Jehee, J., & Murre, J. (2007). Neural models that convince: Model hierarchies and other strategies to bridge the gap between behavior and the brain. *Philosophical Psychology*, 20(6), 749–772.
- Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, 72(2), 1024–1027.
- Nagumo, J., Arimoto, S., & Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10), 2061–2070.
- Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, 65(3), 151–166.
- O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669), 452–454.
- O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, 1104(1), 35–53.
- Ogawa, S., Lee, T.-M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24), 9868–9872.
- Øksendal, B. (2003). *Stochastic differential equations*. New York, NY: Springer.
- O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences*, 2(11), 455–462.
- O'Reilly, R. C., Munakata, Y., Frank, M., Hazy, T., et al. (2012). *Computational cognitive neuroscience*. Mainz, Germany: PediaPress.
- Paul, E. J., & Ashby, F. G. (2013). A neurocomputational theory of how explicit learning bootstraps early procedural learning. *Frontiers in Computational Neuroscience*, 7, Article 177.
- Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, 113(1), 57–83.
- Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3), 472–491.
- Rall, W. (1967). Distinguishing theoretical synaptic potentials computed for different soma-dendritic distributions of synaptic input. *Journal of Neurophysiology*, 30(5), 1138–1168.
- Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, 20(2), 262–270.
- Ricciardi, L. M., & Sacerdote, L. (1979). The Ornstein–Uhlenbeck process as a model for neuronal activity. *Biological Cybernetics*, 35(1), 1–9.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.

- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*, Vol. 1. Cambridge, MA: MIT Press.
- Sandrini, M., Umiltà, C., & Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: A new synthesis of methodological issues. *Neuroscience & Biobehavioral Reviews*, 35(3), 516–536.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80(1), 1–27.
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 26, 241–263.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Seamans, J. K., & Robbins, T. W. (2010). Dopamine modulation of the prefrontal cortex and cognitive function. In K. A. Neve (ed.), *The dopamine receptors*, second edition (pp. 373–398). New York, NY: Springer.
- Segev, I., Fleshman, J. W., & Burke, R. E. (1989). Compartmental models of complex neurons. In C. Koch & I. Segev (eds.), *Methods in neuronal modeling* (pp. 63–96). Cambridge, MA: MIT Press.
- Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4), 1916–1936.
- Sjöström, P. J., Rancz, E. A., Roth, A., & Häusser, M. (2008). Dendritic excitability and synaptic plasticity. *Physiological Reviews*, 88(2), 769–840.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168.
- Stanton, P. K., Bramham, C., & Scharfman, H. E. (2006). *Synaptic plasticity and transsynaptic signaling*. New York, NY: Springer Science & Business Media.
- Strogatz, S. H. (2014). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Reading, MA: Addison-Wesley.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
- Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *The Journal of Neuroscience*, 23(32), 10402–10410.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York, NY: Cambridge University Press.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592.
- Valentin, V. V., Maddox, W. T., & Ashby, F. G. (2016). Dopamine dependence in aggregate feedback learning: A computational cognitive neuroscience approach. *Brain & Cognition*, 109, 1–18.
- Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, 7(2), 208–256.
- Verwey, W. B., Lammens, R., & van Honk, J. (2002). On the role of the SMA in the discrete sequence production task: A TMS study. *Neuropsychologia*, 40(8), 1268–1276.
- Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. *Neuron*, 60(2), 215–234.

- Wiggins, S. (2003). *Introduction to applied nonlinear dynamical systems and chaos, Vol. 2.* New York, NY: Springer Science & Business Media.
- Willingham, D. B., Wells, L. A., Farrell, J. M., & Stemwedel, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, 28(3), 366–375.
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1), 1–24.
- Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of nervous tissue. *Kybernetik*, 13(2), 55–80.
- Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G. C., Urakubo, H., Ishii, S., & Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, 345(6204), 1616–1620.
- Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A., & Poo, M.-M. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395(6697), 37–44.

7 Discovering Aggregation Properties via Voting

Donald G. Saari

7.1	Aggregations	271
7.2	Limitations of Paired Comparisons	274
7.2.1	Reductionist Approach with N Alternatives	274
7.2.2	Quasi-transitivity	278
7.3	Source of Voting Problems	281
7.3.1	Computing Tallies	281
7.3.2	Consequences	282
7.3.3	A Basis for Outcomes	284
7.3.4	Applications	290
7.4	Paired Comparisons: An Analysis	296
7.4.1	Paired Comparison Outcomes	297
7.4.2	Transitivity	300
7.4.3	Transitivity vs. Strong Transitivity	302
7.4.4	Seeking Perversities	303
7.4.5	Basis in Profile Space	305
7.5	Selective Consequences	309
7.5.1	Voting: How Can These Results Be Used?	309
7.5.2	Resolving Arrow's Theorem	313
7.5.3	Path-dependency and Quantum Thinking	314
7.5.4	Consistency in Ratio-scaled Comparisons	316
7.6	Summary	319
	References	320

7.1 Aggregations

Aggregation approaches, ranging from statistics to resource allocations, are central to the social and behavioral sciences. Probably the most widely used are voting methods where, hopefully, individual preferences are combined to capture the societal intent. But casting doubt as to whether an election outcome accurately reflects the voters' aggregate wishes are examples where, with the same data, different voting rules have conflicting conclusions.

If, for instance, the preferences for 16 voters over candidates A , B , C are

Number	Ranking	Number	Ranking
3	$A > C > B$	4	$A > B > C$
3	$B > C > A$	6	$C > B > A$,

(7.1)

then A wins with the vote-for-one rule (plurality vote), B wins with the vote-for-two rule, and C wins with majority votes over the three pairs. *Each candidate* “wins” with an accepted election method, so who is the voters’ true aggregate choice?

Adding to the discomfort are the many paradoxical examples (e.g., Theorem 7.5 below) where, for instance, the plurality vote is $A > B > C > D$, but if D withdraws, the voters’ sincere plurality vote is $C > B > A$; if C , rather than D , withdrew, the plurality outcome is $D > B > A$; and if B , rather than C or D , withdrew, the plurality outcome is $D > C > A$. A is the plurality winner over all four candidates, but she is plurality bottom-ranked over all triplets. Is she, or is she not, the voter’s preferred choice?

What amplifies the importance of “voting” is its role as a prototype for general aggregation methods. For this reason, expect the complexities exhibited by voting rules to affect other aggregations; e.g., they plague non-parametric statistics (Section 7.4.1). Mimicking the above comment about how a four-alternative voting outcome can conflict with those of three alternatives is how, in a pure-exchange economy, the four-commodity aggregate excess demand function can differ as radically as desired from the excess demand for any three-commodity subset (Saari, 1992a). However, an encouraging possibility is that recent results in voting theory explain what causes all possible standard voting difficulties. These arguments may transfer to clarify peculiarities from other areas, including the cognitive sciences.

Tversky (1969), for instance, showed that individuals can make cyclic decisions in multi-attribute settings. With lotteries (a, p) , where p is the probability of receiving \$a, and $(1 - p)$ is the likelihood of getting \$0, he found with paired comparisons that many people exhibited the cyclic outcomes

$$(5.00, \frac{7}{24}) \succ (4.75, \frac{8}{24}) \succ (4.50, \frac{9}{24}) \succ (4.25, \frac{10}{24}) \succ (4.00, \frac{11}{24}) \succ (5.00, \frac{7}{24}) \quad (7.2)$$

over the five options.

What causes this cycle? It is reasonable to expect that this intransitivity reflects particular features in the presentation. This is captured, for instance, with four identical acts

	1	2	3	4
x_1	10,000	9,000	8,000	7,000
x_2	9,000	8,000	7,000	10,000
x_3	8,000	7,000	10,000	9,000
x_4	7,000	10,000	9,000	8,000

(7.3)

where each state has the same 25% chance for success; the tendency for the $x_1 \succ x_2 \succ x_3 \succ x_4$ ranking may reflect an inclination to emphasize the first entry.

A full description of Equation (7.2) requires ruling out other explanations. Are there are subtle aspects that contribute to a cyclic decision behavior? After all, it is reasonable to assume that each individual has a “more is better” transitive ranking for each attribute as in

$$\begin{aligned} \text{Probability ranking: } & \frac{11}{24} \succ \frac{10}{24} \succ \frac{9}{24} \succ \frac{8}{24} \succ \frac{7}{24} \\ \text{Cash ranking: } & 5.00 \succ 4.75 \succ 4.50 \succ 4.25 \succ 4.00. \end{aligned} \quad (7.4)$$

If A, B, C, D, E represent the lotteries in the Equation (7.2) order (so $A = (5.00, \frac{7}{24})$ and $E = (4.00, \frac{11}{24})$), what are *all* ways to aggregate the transitive information

Attribute	Ranking	
Probability	$E \succ D \succ C \succ B \succ A$	
Cash	$A \succ B \succ C \succ D \succ E$	

(7.5)

into the observed $A \succ B, B \succ C, C \succ D, D \succ E, E \succ A$ cyclic decisions? (This example is briefly discussed after Theorem 7.14.) A deeper question (Section 7.2) is whether we should be surprised with intransitive outcomes when individuals use paired comparisons.

What motivates these comments is that, as known since Condorcet’s work in 1785, even seemingly well-behaved voting rules can convert transitive inputs into cyclic outcomes. For a five-option example, suppose there are three voters with the transitive preferences

$$A \succ B \succ C \succ D \succ E, \quad B \succ C \succ D \succ E \succ A, \quad C \succ D \succ E \succ A \succ B. \quad (7.6)$$

Everyone prefers $C \succ D \succ E$, yet pairwise voting defines the majority vote cycle

$$A \succ B, \quad B \succ C, \quad C \succ D, \quad D \succ E, \quad E \succ A.$$

A three-attribute individual-decision example mimicking Equation (7.6) can be expected to also generate a cycle. The question is whether other data structures have this property.

This question reflects a central goal in voting theory, which is to reach beyond designing examples to explain all behaviors. A successful approach (outlined here) identifies *all possible ways* to create cycles and other paradoxical outcomes (Saari, 1999, 2000a, 2000b, 2008). As discovered, different, subtle aspects of the data (e.g., profiles; lists of the voter preferences) force different rules to have different outcomes. To create Example 7.1, for instance, I combined two different data structures, where one *never* affects pairwise rankings but forces differences in the plurality and vote-for-two outcomes, while the second determined the pairwise outcomes. If data structures can explain all voting differences, then, with voting’s exemplar status for aggregation rules, it is reasonable to explore whether similar structures contribute to our understanding of individual decisions and other aggregation phenomena.

As paired comparisons are widely used in the cognitive sciences, my emphasis is on pairwise voting (which now is well understood (Sections 7.3 and 7.4)). However, paired comparisons include controversies and incomplete analyses. For instance, in response to concerns about Saaty's analytic hierarchy process (AHP) (Saaty, 1980), Crawford and Williams (1985) and Crawford (1987) proposed a “geometric mean” approach as an alternative for ratio-scaled comparisons. Both methods were compared with a Monte Carlo study, but I know of no theoretical explanation of the two method’s relative strengths and weaknesses. Fortunately, the voting structures described here permit doing so; the analysis is in Section 7.5.4.

Other topics briefly described (in terms of voting structures) include Luce’s axioms (Luce, 1959) for individual decisions and an alternative for the recent interest in quantum thinking.

7.2 Limitations of Paired Comparisons

The names of the N alternatives are given by capital letters such as A, B, \dots , or A_1, A_2, \dots, A_N . In addition to the “strictly preferred” symbol “ \succ ,” tied situations are represented by “ \sim .” A standard goal is to have transitive outcomes (where $A_i \succ A_j$ and $A_j \succ A_k$ require $A_i \succ A_k$), but Equation (7.6) proves that this objective can fail with majority votes over pairs. A natural response is to design approaches that guarantee transitivity. However, as shown next, this cannot be done with standard assumptions.

Indeed, a major conclusion in voting theory is “Arrow’s impossibility theorem” (Arrow, 1951), which commonly is described as asserting that a fair voting method cannot exist with three or more candidates. What is not fully appreciated is that Arrow’s result applies to general aggregation methods coming from most disciplines. To show why this is so, the slight extension given here (from Saari, 1995a) emphasizes transitive inputs that are independent of whether they represent voter preferences or a given data structure.

7.2.1 Reductionist Approach with N Alternatives

Creating an appropriate aggregation rule can be difficult. To simplify the design process, adopt the reductionist methodology by dividing the analysis into simpler, more tractable parts of paired comparisons. For each pair, design a ranking rule. (Different pairs can have different rules.) The following (Saari, 1995a) makes this precise.

- (1) *Inputs:* Each of the $n \geq 2$ agents (i.e., voters, attributes, sources of data, etc.) has a complete transitive ranking of the $N \geq 3$ alternatives; no restrictions are imposed on the rankings.
- (2) *Outputs:* This information is to be combined into a complete transitive ranking.
- (3) *Reductionist approach (paired comparisons):* For each pair, define a rule that uses information only about the relative ranking of these two alternatives.

Namely, for two different data sets \mathbf{d}_1 and \mathbf{d}_2 , if each agent has the same ranking for the pair in \mathbf{d}_1 and \mathbf{d}_2 , then the pair's outcome ranking is the same for \mathbf{d}_1 and \mathbf{d}_2 .

- (4) *Involvement:* A pair's rule cannot have a constant outcome. Namely, each of at least two of the three ways to rank a pair must arise with some data set.
- (5) *Aggregation:* All pairwise outcomes cannot always be determined by the same agent.

The first two conditions are standard transitivity requirements for the domain and range. The third (reductionist approach) specifies the methodology of using non-constant valued (Condition 4) paired comparisons. To be an aggregation (Condition 5), there must be some situation where inputs from more than one agent, or data source, are involved. To illustrate Condition 5's flexibility, with $N = 3$, one agent, perhaps because of expertise, could always determine the outcome of one, or even two of the pairs as long as there is at least one setting where another agent influences the third pair's outcome.

Surprisingly, these highly flexible conditions cannot always be satisfied! No matter how cleverly paired comparison rules are designed to satisfy Conditions 1, 3–5, they will experience intransitivities! (This conclusion must be expected with unruly data, so treat Condition 1 as asserting that the negative assertion holds *even* with well-behaved data.)

Theorem 7.1 *It is impossible to create paired comparisons decision rules that always satisfy the above five conditions.*

To explore what Theorem 7.1 might mean about multi-attribute individual decision processes, the individual is the aggregation method and the $n \geq 2$ attributes serve as the $n \geq 2$ agents. The individual's decision rules can vary with the pair; e.g., with lotteries of the Equation (7.2) type, decisions for some pairs might compare expected values, for others the reward size, or probability value might matter. According to Theorem 7.1, *it does not matter what rules are used* (Condition 3). Even in an idealized setting where, with each attribute, an individual has a complete transitive ranking (Condition 1) and not all decisions are based on a single attribute (Condition 5), intransitive outcomes must be anticipated (Condition 2 and Theorem 7.1) And so intransitive outcomes in individual decisions should not be that surprising.

Arrow's theorem is a special case of Theorem 7.1: for Arrow, agents (the data sources) are voters; involvement (Condition 4) is replaced with the more restrictive unanimity requirement (weak Pareto) where, for any pair, if all agents have the same strict ranking of the pair, then that is the pair's outcome. Arrow's "independence of irrelevant alternatives" (IIA) is stated differently, but agrees with the reductionist Condition 3; his "no dictator" assumption is a special case of Condition 5. However, rather than the "no voting rule is fair" assertion, a more accurate interpretation of Theorem 7.1 and Arrow's theorem is that it offers

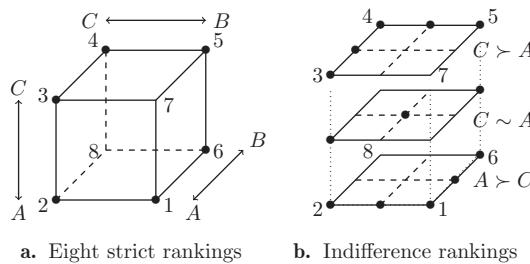


Figure 7.1 *Geometry of rankings.*

a warning: *No matter how cleverly the reductionist approach with the paired comparison methodology is used, it must experience difficulties.*

Arrow's seminal contribution, which has strongly influenced voting theory, is often treated as a mystery. No longer should this be the case, because the following proof is not difficult; it essentially reflects the geometric arrangement of intransitive rankings relative to the transitive ones. (This geometric approach was developed in Saari 1995a, 2001b.) For $N = 3$, this arrangement is illustrated with the Figure 7.1a cube.

Let points $-1, 0, 1$ on the x axis represent, respectively, $B > A, A \sim B, A > B$. Similarly, $-1, 0, 1$ on the y -axis represents $C > B, B \sim C, B > C$, and on the z axis $A > C, A \sim C, C > A$. A cube has eight vertices; in Figure 7.1a, the six transitive rankings are indicated with bullets; the remaining two vertices, 7 and 8, correspond to cyclic rankings. The vertices and their assigned numerical names are

Name	Ranking	Name	Ranking
1	$A > B > C$	2	$A > C > B$
3	$C > A > B$	4	$C > B > A$
5	$B > C > A$	6	$B > A > C$
7	$A > B, B > C, C > A$	8	$B > A, A > C, C > B$

The geometry, then, requires an $A > B$ outcome to be on the front face with vertices $\{1, 2, 3, 7\}$. If $A > B$ is accompanied with $B > C$, the outcome is on the intersection of the front and right faces, which is the front-right edge with vertices $\{1, 7\}$. Coupled with the $A > C$ outcome, the conclusion is the intersection of the bottom face with the front-right edge, or vertex 1 with the $A > B > C$ ranking.

Central for the proof is that *each of the cube's six faces* has a vertex representing cyclic rankings (vertex 7 or 8); e.g., a pair's strict outcome defines a face with a cyclic vertex, the full outcome is determined by the choices of the face's two pairs of edges. If one agent controls the choice for one pair, and another agent independently controls the other pair, then, obviously, there is a combination that forces a cyclic outcome.

The $C > A$ outcome, for instance, defines the Figure 7.1a top-face with vertices $\{3, 4, 5, 7\}$. The remaining pairs are $\{A, B\}$ and $\{B, C\}$; different strict $\{A, B\}$ outcomes correspond to selecting either the top-front ($A > B$) or top-back ($B > A$)

edge, while different $\{B, C\}$ choices correspond to selecting either the top-right ($B \succ C$) or top-left ($C \succ B$) edge. If the edge selections for the two pairs can be made independent of each other, then selecting top-front and top-right forces the outcome to be the cyclic vertex 7.

Namely, it is the positioning of intransitive vertices relative to the transitive ones combined with the ability to independently change rankings for the other pairs that forces intransitive outcomes. The only requirement is the following property:

With any ranking of one pair, settings exist where at least two rankings for each of the two other pairs can be selected independent of each other.

(7.8)

A similar structure holds should ties be included. As shown in Figure 7.1b, adding indifference (i.e., some x, y , and/or z component equals 0) creates a grid structure with nine points; *only* the points with bullets are transitive! Displayed from top down are the $z = 1$ face ($C \succ A$), the $z = 0$ face ($C \sim A$), and the $z = -1$ face ($A \succ C$). Three alternatives define the 13 transitive rankings represented by bullets. Five are on the top, five are on the bottom, and three are on the $z = 0$ face. The midpoint of the top-front edge, for instance, does not have a bullet because it represents the $C \succ A$ ($z = 1$ face), $A \succ B$ (top-front edge), $C \sim B$ ($y = 0$ line) intransitive rankings. Similarly for the $z = 0$ plane with $C \sim A$, its back-left vertex does not have a bullet because it represents the $C \succ B$ (left edge), $B \succ A$ (back edge), and $C \sim A$ (the $z = 0$ face) intransitive rankings.

For any face and any two choices for each of the two remaining pairs, some combination of the lines crosses an intransitive point. Thus, once one pair's outcome is determined, the ability to independently select at least two choices for each of two remaining pairs (Property 7.8) can force intransitive outcomes. The proof of Theorem 7.1 shows that this happens.

Proof The proof is given for $N = 3$, where Figure 7.1a now is used to represent agents' preferences. (Simple modifications, which essentially fix the other variables, extend the proof to $N \geq 3$.) All that is needed is to show that Property 7.8 holds.

Assume Theorem 7.1 is false; i.e., paired comparison rules exist that satisfy Conditions 1–5. Thus, at least two agents, **a** and **b**, can, in appropriate settings, change the outcome of different pairs. So (without loss of generality) a situation exists (i.e., a data set or profile) where if **a** changes $\{A, B\}$ preferences, the $\{A, B\}$ outcome changes (involvement). Of course, this change may require each of the other agents to have a specific $A \succ B$, or $B \succ A$ ranking. Similarly, a situation exists where when **b** changes $\{B, C\}$ preferences, the $\{B, C\}$ outcome changes; again, each of the other agents may need to have a specific $B \succ C$, or $C \succ B$ ranking. Let each agent other than **a** and **b** have a transitive ranking consistent with their two assigned choices. (This is one of the few places where transitivity plays a strong role; i.e., rankings of any two pairs can be combined in a transitive manner.)

The condition to empower agent **a** may require agent **b** to have a specific $\{A, B\}$ ranking; i.e., either **b**'s rankings (vertices) come from the front or back face. If it is

the front face ($A > B$), let **b**'s choices be between the vertices on the bottom edge, which fixes **b**'s $\{A, B\}$ and $\{A, C\}$ rankings. If **b** must have a back face ranking (i.e., $B > A$), **b**'s two choices are from the top edge (which, again, fixes **b**'s $\{A, B\}$ and $\{A, C\}$ rankings). Similarly, for **b** to have the power to decide, **a** may need a specific $\{B, C\}$ ranking; if it defines the right face (with vertices $\{1, 5, 6, 7\}$), let the two choices be between the vertices on the face's bottom edge, which have the same $\{B, C\}$ and $\{A, C\}$ rankings. If it is the left face, use the vertices from the top edge which have the same $\{B, C\}$ and $\{A, C\}$ rankings.

By construction, *each agent has a fixed $\{A, C\}$ preference ranking*, which requires (reductionist; paired comparison) a fixed $\{A, C\}$ outcome. With these choices, independent of what the other agent does, **a** can alter the $\{A, B\}$ outcome and **b** can alter $\{B, C\}$. This satisfies Property 7.8, and proves the theorem. \square

The impossibility conclusion can be avoided by allowing the outcomes to depend on a single voter's preferences – a dictator. This is because, by relying on a single agent, what happens with one pair restricts what can happen with other pairs. Thus the Equation (7.8) property, where outcomes for pairs can be independently selected, cannot be satisfied.

If paired comparisons suffer problems, perhaps we could use triplets, or quadruples, or ... No; a slight modification of the above statement and proof shows that the same difficulty persists. With triplets and $N = 4$ alternatives, for instance, a geometric structure (in \mathbb{R}^8) can be created with the six transitive rankings for each of the four triplets where the convex hull includes even more intransitives.

7.2.2 Quasi-transitivity

It is easy to design paired comparison rules that satisfy Conditions 1, 3–5 by following the lead of what is often done in practice where precision is lacking. Namely, replace the requirement (Condition 2) of having transitive outcomes with the weaker acyclicity condition that requires no cycles. Let me illustrate with a highly prejudiced rule where, with $1 \leq i < j \leq N$, if A_i beats A_j in a majority vote (notice, $i < j$), the ranking is $A_i > A_j$; otherwise the ranking is $A_i \sim A_j$. (So, if A_j beats A_i where $j > i$, the outcome is listed as a tie.) This condition converts a $A_1 > A_2, A_2 > A_3, A_3 > A_1$ majority vote cycle into an intransitive, but acyclic $A_1 > A_2, A_2 > A_3, A_1 \sim A_3$ triplet.

This silly rule has obvious flaws, but it suggests the value of exploring whether the Theorem 7.1 conclusion can be avoided by replacing transitivity of paired comparison outcomes with more relaxed requirements. It is obvious what must be done; find a condition that adds enough new designated points to Figure 7.1a,b to prevent Property 7.8 from realizing a non-identified point. A possible approach is “quasi-transitivity,” which is defined below.

But first, to geometrically describe transitivity, let points indicating the level of support (perhaps they are election tallies) for candidates $\{A, B, C\}$ be given, respectively, by $(X, Y, Z) \in \mathbb{R}_+^3$ (the “+” subscript indicates that all entries are

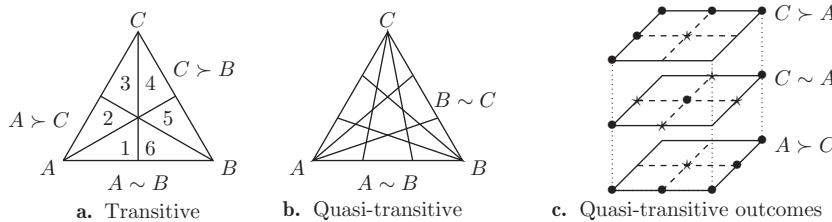


Figure 7.2 Generalization from transitive to quasi-transitive.

non-negative). Normalize these values (i.e., divide each term by $X + Y + Z$, e.g., $x = \frac{X}{X+Y+Z}$) to create a point on the unit simplex (represented by the Figure 7.2a equilateral triangle)

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}_+^3 \mid x + y + z = 1, x, y, z \geq 0\}.$$

This equilateral triangle (Figure 7.2a) will be divided into regions that represent the 13 different $(x, y, z) \in \mathcal{S}$ rankings. All points on the vertical line passing through the C vertex are equal distance from the A and B vertices, so they represent the tied $A \sim B$ outcome. Points in the right triangle to the left of this line are closer to the A vertex, so they have the $A > B$ ranking; points in the other right triangle have the $B > A$ ranking. Similarly, the line passing from the midpoint of the $\{B, C\}$ edge to vertex A represents $B \sim C$, and the line from the midpoint of the $\{A, C\}$ edge to vertex B represents $A \sim C$.

In this way, each of the six Figure 7.2 small triangles corresponds to a strict (i.e., no ties) transitive ranking; e.g., all points in the Figure 7.2a triangle with a “1” are closest to A , next closest to B , and farthest from C , so the ranking is $A > B > C$. The number in each small triangle is the region’s Equation (7.7) ranking name. To further illustrate, the Figure 7.2a large $C > B$ right triangle consists of all five rankings involving $C > B$; the strict rankings are given by triangles 2, 3, 4, the lower line segment has a tie to be $A \sim C > B$, and the upper segment represents $C > B \sim A$.

It is interesting how this geometry captures the transitivity conditions. For instance, the Figure 7.2a intersection of the $A > C$ and $C > B$ right triangle is triangle 2; by being in the $A > B$ triangle, this intersection captures the “ $A > C$ and $C > B$ requires $A > B$ ” transitivity condition. Similarly, the intersection of the $A \sim B$ vertical line with the $C > B$ right triangle is the short vertical $C > B \sim A$ segment manifesting the transitivity “ $A \sim B$ and $C > B$ requires $C > A$ ” condition.

Quasi-transitivity retains the transitive relationships for strict rankings while adding new ones with indifference; e.g., it allows “ $A \sim B$ and $B \sim C$ to admit $A > C$,” reflecting the familiar “Ann is essentially as tall as Barb, and Barb is essentially as tall as Connie, but Ann is taller than Connie.” All quasi-transitive conditions are captured by Figure 7.2b, which enlarges the Figure 7.2a indifference lines into wedges. As with Figure 7.2a, the Figure 7.2b intersection of the

$C \succ B$ and $A \succ C$ triangles is the small $A \succ B$ triangle, so quasi-transitivity also requires $A \succ C$ and $C \succ B$ to be accompanied by $A \succ B$.

For new conditions, with transitivity, $C \succ B$ and $B \sim A$ requires $C \succ A$ for the transitive $C \succ B \sim A$. In contrast, the Figure 7.2b intersection of these two regions consists of a quadrilateral and small triangle. The quadrilateral (with C as a vertex) is the transitive “ $C \succ B$ and $A \sim B$ with $C \succ A$ ” triplet. The smaller triangle is in the $A \sim C$ region, so it defines the quasi-transitive (but intransitive) “ $C \succ B$ and $A \sim B$ can admit $A \sim C$ ” rankings. Similarly, with transitivity, $A \sim B, B \sim C$ requires (the intersection of the Figure 7.1a vertical and upward-slanting indifference lines) $A \sim C$. With quasi-transitivity and Figure 7.2b, the intersection of the two wedges has *three* regions, a hexagon and two small triangles, so $A \sim B, B \sim C$ can be accompanied with any of $A \succ C, A \sim C$, or $C \succ A$.

With three alternatives, transitivity has 13 sets of rankings depicted by the Figure 7.1b bullets. However (Figure 7.2b,c), quasi-transitivity admits 19 sets; the 13 transitive ones and six new options. All 19 are depicted in Figure 7.2b. In the Figure 7.2c cube, the transitive rankings are denoted by bullets and the new additions by stars. The issue is whether there are enough stars and bullets to convert Theorem 7.1 to a positive conclusion; that is, are there enough new points to counter the Property 7.8 possibility of selecting a non-designated point. The answer is not immediate; e.g., the rule defined in this section’s first paragraph has acyclic, but not necessarily quasi-transitive outcomes.

To simplify the argument, replace the “involvement” with weak Pareto, where if all agents have the same ranking for a particular pair, then that is the pair’s outcome. Examining the geometry, it follows that the only wiggle room is if agents disagree about a pair, then the outcome must be a tie. This observation motivates the following:

Definition 7.1 An oligarchy is a set of k agents, $2 \leq k \leq n$, where when all are in agreement over a pair, that is the pair’s outcome. However, if even one agent disagrees, the pair’s outcome is a tie. (A one-person oligarchy is a dictator.)

It is not difficult to show that if transitivity is replaced with quasi-transitive outcomes, then an oligarchy creates a rule. Gibbard (1969) proved it is the only approach, which, for users of paired comparisons, means that problems remain; expect difficulties. (Ways to circumvent these concerns are described in Section 7.5.2.)

Theorem 7.2 (Gibbard, 1969) *For $N \geq 3$ alternatives and $n \geq 2$ agents, there does not exist a preference aggregation rule where agents have complete transitive preference rankings (without restriction) and the rule satisfies quasi-transitivity, completeness of social preferences, the weak Pareto principle, IIA, and non-oligarchy.*

Left for the reader is to determine what happens should weak Pareto be replaced with involvement. Also, relative to interests from the behavioral sciences, what happens if agents (e.g., sources of data) have quasi-transitive preferences?

7.3 Source of Voting Problems

Let's move away from negative assertions to discover everything that can happen with standard voting rules. The approach I developed (e.g., Saari, 1999) is described here with $N = 3$ alternatives. Readers primarily interested in an explanation of what causes all $N \geq 3$ paired comparison voting problems can jump to Section 7.4.

7.3.1 Computing Tallies

An interesting fact (developed next) is that all three-alternative difficulties reflect the geometric symmetries of the Figure 7.2a triangle. A way to appreciate these symmetries is to use them to simplify the computation of tallies. Adding tallies is elementary; what is annoying is sifting the ballots to see who is ranked where. Here, the geometry saves effort.

The approach uses what I call a “geometric profile representation.” Illustrating with the Equation (7.1) profile, place the number of voters with a particular preference ranking in the appropriate region. Four voters, for instance, prefer $A > B > C$, so place “4” in the $A > B > C$ small Figure 7.3a triangle (which is region 1 in Figure 7.2a). Similarly, nobody has the $C > A > B$ ranking, so zero is placed in this Figure 7.3a region (near the C vertex).

The reason this geometric sorting simplifies tallying pairwise majority votes is that all numbers in the $A > B$ right triangle (the shaded Figure 7.3b region) have $A > B$ preferences, so A 's tally is the sum of the region's entries. This $0 + 3 + 4 = 7$ tally in Figure 7.3a is placed under the $A-B$ edge closer to the A vertex. Similarly, the sum of entries in the right triangle to the right of the $A \sim B$ vertical line, $6 + 3 + 0 = 9$, is B 's tally, so 9 is placed near B under the $A-B$ edge. The tallies for the remaining two pairs are similarly computed and listed near their respective triangle edge. According to these tallies, the outcomes are $C > A$, $C > B$, $B > A$, each by 9:7.

The plurality and vote-for-two rules are special cases of positional methods.

Definition 7.2 A three-alternative positional rule is where specific weights, (w_1, w_2, w_3) , are used to tally ballots where w_j points is assigned to the j th positioned alternative; the alternatives are ranked according to the sum of received weights. The weights satisfy $w_1 > 0$ and $w_1 \geq w_2 \geq w_3 = 0$.

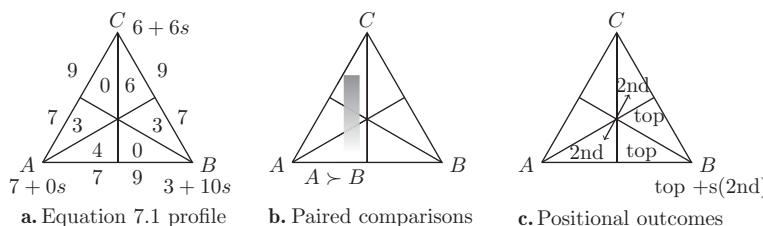


Figure 7.3 Computing tallies.

A normalized positional method is where each weight is divided by w_1 to define

$$\mathbf{w}_s = (1, s = \frac{w_2}{w_1}, 0). \quad (7.9)$$

As examples, the plurality vote has the weights $(1, 0, 0)$ where $s = 0$ means that second-positioned candidates get zero points. In contrast, the vote-for-two rule has weights $(1, 1, 0)$ where a ballot's second-positioned candidate receives the same number of points, $s = 1$, as a top-positioned candidate. Important for what follows is the Borda count defined by $(2, 1, 0)$, where its normalized version $(1, \frac{1}{2}, 0)$ assigns a half a point for second-positioned alternatives. Another choice might be $(7, 2, 0)$ with the normalized form of $(1, \frac{2}{7}, 0)$, where a second positioned alternative on a ballot receives $\frac{2}{7}$ points.

Geometry also sorts a profile's entries (Figure 7.3c) to simplify the tallying of positional methods. Illustrating with alternative B , its plurality vote (\mathbf{w}_0) is the number of ballots with B top-positioned, which is the sum of numbers in the two regions with B as a vertex; in Figure 7.3a, this is $3 + 0 = 3$. The \mathbf{w}_s tally, where $s \neq 0$, is the plurality vote plus s times the number of ballots with B second-positioned. This second-positioned status is given by the two regions adjacent to the top-positioned state (see Figure 7.3c), so B 's tally for $(1, s, 0)$ is $3 + s(4 + 6) = 3 + 10s$, which is posted by the B vertex. The full Figure 7.3a election outcome, with values listed near the appropriate vertex, is

$$A : B : C = (7 + 0s, 3 + 10s, 6 + 6s). \quad (7.10)$$

7.3.2 Consequences

According to Equation (7.10), the plurality outcome ($s = 0$) is $A > C > B$ with 7:6:3 tally, the vote-for-two outcome ($s = 1$) is the *reversed* $B > C > A$ with 13:12:7 tally, and the (normalized) Borda outcome is $C > B > A$ with 9:8:7 tally. (The normalized Borda weights are obtained by dividing the usual weights by two, so the standard Borda tally is twice this or $C > B > A$ with a 18:16:14 tally.) Notice; each alternative is the “winner” with a specified positional method. Even more; points in the Figure 7.2a triangle can be treated as election tallies. By doing so and plotting Equation (7.10), it becomes clear that the resulting line (the “procedure line”; Saari, 1995a) crosses four open triangles and three line segments. Thus this simple Equation (7.1) profile admits *seven different positional outcomes* where four are strict and three have ties!

As demonstrated, a single profile can cause conflicting outcomes, which causes us to wonder whether such behavior is rare or likely. The result is that they are to be expected; a former graduate student, Monica Tataru, and I (Saari & Tataru, 1999) used central limit arguments to prove that with about 69% of close elections, a profile has multiple positional outcomes! But, by not being searched for, the behavior may not have been recognized in practice.

Theorem 7.3 (Saari & Tataru, 1999) *With standard central limit assumptions,¹ in the limit as the number of voters, n , approaches infinity, the probability a profile has precisely k different election outcomes obtained by varying the choice of the positional voting method is*

k	Probability	k	Probability
1	0.31	3	0.44
5	0.19	7	0.06

(7.11)

The limit probability of an even number of different outcomes is zero; this is because either the plurality or the vote-for-two outcome has a tied outcome.

And so (Theorem 7.3), it is more than twice as likely for a profile to have multiple three-alternative election outcomes than consistency. As should be expected, problems seriously escalate with an added number of alternatives. Here the procedure line morphs into a procedure hull (with vertices given by the “vote-for- k -candidates” outcomes) where the number of possible outcomes admitted by a profile rapidly grow.

Theorem 7.4 (Saari, 1992b) *For $N \geq 2$ and any integer k , $1 \leq k \leq (N-1)[(N-1)!]$, a profile exists (actually, an open set of them) with precisely k different strict election outcomes that arise by changing the positional method.*

With $N = 10$ alternatives, then, a profile exists with $9(9!) = 3,265,920$ different election rankings (without ties) that emerge by changing the positional method! Indeed, each alternative could be in first place with some methods, second with others, and last in other methods’ outcomes. So, who is the voters’ true choice? To underscore the concern, the 2016 Republican presidential primaries started with $N = 17$ candidates; 17 candidates admits an open set of profiles where each generates hundreds of trillions (i.e., $16(16!) = 334,764,638,208,000$) of different positional outcomes.² Bothersome is that, as now known, the voting method actually used, the plurality vote, is one of the least reliable indicators of voter intent. Did this voting rule affect the final choice of the Republican primary? More generally, which method has the most reliable outcomes?

Voting outcomes can be surprisingly bad! Theorem 7.3, the above discussion, and the following result indicate why I find it difficult to trust voting results.

Theorem 7.5 (Saari, 1989) *For $N \geq 3$ alternatives,*

- *select a complete transitive ranking for the set of all N alternatives,*
- *for each integer j , $2 < j < N$, and each of the $\binom{N}{j}$ subsets of j alternatives, select, in any desired manner (e.g., even randomly), a complete transitive ranking,*
- *for each of the $\binom{N}{2}$ pairs, select a ranking.*

1 As the number of voters tends to infinity, the probability choices of preferences are asymptotically identical and independently distributed with an asymptotic common finite and positive variance and an asymptotic mean with an equal distribution of voters of each type.

2 Only about six or seven candidates had any chance of winning, so an actual profile for the 2016 primaries probably would admit only about 500–600 different election outcomes.

There exists a profile where each set's plurality election outcome and each pair's majority vote ranking is as selected.

What a mess! According to Theorem 7.5, any imagined (or feared) paradoxical behavior actually exists! The full situation is much worse; I further proved (Saari, 1989) that this negative conclusion, where *anything* can happen, holds for almost all positional voting rules. Only the Borda count (tally a ballot by assigning $N - j$ points to a j th positioned candidate) provides consistency and avoids huge numbers of difficulties.³ Of particular concern is that these difficulties plague other aggregation methods, which is captured by results in Saari (1995b) (where this behavior is connected with “chaotic dynamics”) showing that similar conclusions affect probability, statistics, etc.

7.3.3 A Basis for Outcomes

So many things can go wrong, but why? While my emphasis is on $N = 3$ alternatives, this approach (of appealing to symmetries) extends to $N \geq 3$.

First, some history. Aggregation approaches involve the three major components

$$F : \{\text{Data Space}\} \rightarrow \{\text{outcome space}\} \quad (7.12)$$

of the mapping F , the range, and the domain. Stimulated by Arrow’s success in using the theoretical method to discover his impossibility theorem (Section 7.2), a focus was placed on exploring properties of mappings. This led to the so-called *axiomatic approach* of using “axioms” to eliminate certain F choices and identify others. An often-asserted rationale for doing so was that, by relying on axioms, “we know what we are getting.” Unfortunately, in the manner the approach was generally used, this comment proved to be false.

While the axiomatic movement uncovered interesting conclusions (some with challenging proofs), most assertions have not survived the test of time. It is fair to say that after over 65 years of effort, this approach provided minimal guidance about what should be done. In part (as described in Saari, 2008, chapter 3), this is because, in general, rather than “axioms” (which *are* fundamental building blocks that identify what “we are getting”) standard “hypotheses” were being used; these conditions so highly restricted the admissible profiles that only specific methods had the specified properties. (In contrast, Luce’s axiom (Section 7.3.4.2) *is* an axiom; to see this, *all* of his major conclusions can be traced back to his axioms.) An extreme example, which captures what was done, are the “axioms”

1. man,
2. living at a specific address

³ Call a list of Theorem 7.5 plurality outcomes – one ranking for each set of alternatives – a “word.” With only seven alternatives, the number of different plurality words is more than 10^{50} times the number of different Borda words, which indicates the significant magnitude of paradoxical settings avoided by the Borda count.

that uniquely identify me but provide no added information. These are not axioms; with them, it is not known what else you are getting.⁴ Incidentally, because “axiomatic” conclusions tend to describe what happens with special choices of profile, they become specific cases of Theorem 7.5 and its generalizations. Thus trillions of other “axiomatic conclusions” have yet to be rediscovered – with only seven alternatives.

An alternative approach is to emphasize the structure of the range, or “outcome space” of Equation (7.12). The goal is to impose conditions on what constitutes an acceptable outcome. Here the mappings F and possible restrictions on the data space are treated as variables to be satisfied. R. D. Luce (1959, 1977; see also Yellott, 1997) successfully initiated this endeavor with his axioms for individual decision-making. His well-received contributions influenced advances in different directions and disciplines. Unfortunately, his work was overlooked by the social choice community. (Luce’s description used probabilities of different rankings, but it is easy to convert probabilities into numbers of voters with particular preference rankings.) A description of how voting theory helps to advance some of Luce’s work is in Section 7.3.4.2.

What remains in Equation (7.12) is the domain. To determine what can happen, I initiated a study of the data space structure for decision approaches and, in particular, voting methods. For voting, there are $N!$ ways to rank N alternatives, so the domain – the space of profiles – can be identified with $\mathbb{R}^{N!}$. Here a vector in $\mathbb{R}^{N!}$ specifies the number of agents with a specific ranking. (With more general settings, such as probabilities or where a voter’s ballot is assigned a weight, an $\mathbb{R}^{N!}$ vector need not have integer components.) The goal, as outlined next, is to find an orthogonal basis for $\mathbb{R}^{N!}$. That is, decompose the data structure into those components that cause different methods to exhibit different behavior.

7.3.3.1 Basis for Three Alternatives

In retrospect, reasons why different voting methods (and aggregation procedures in general) can have conflicting outcomes become clear from the geometric profile representation (Figure 7.3a) and tallying methods (Figure 7.3b,c). Illustrating with alternative A and using the Figure 7.2a names for ranking regions (also Equation (7.7)):

- A ’s plurality tally is the sum of entries in regions 1 and 2,
- A ’s majority vote against B is the sum of entries in regions 1, 2, and 3, which means that region 3 values can force this tally to differ from A ’s plurality tally.

⁴ One “axiomatic” conclusion, for instance, asserts that the only method satisfying some technical assumptions and that always elects a majority winner is the plurality vote. Another “axiomatic” conclusion proves that the only method satisfying the same technical assumptions but which can elect a candidate who is bottom-ranked by one less than $\frac{N-1}{N}$ of all voters also is the plurality vote. Neither set of “axioms” identifies what we are getting; neither can predict the other conclusion. Rather than axioms, they are special properties that identify isolated profiles with which only the plurality vote has these properties.

- A 's vote-for-two tally is the sum of entries in regions 1, 2, 3, and 6. Thus, region 6 values can force this tally to differ from A 's vote in an $\{A, B\}$ election and entries in regions 3 and 6 can force this value to differ from A 's plurality tally.

All methods have full access to all of the data, but different approaches ignore aspects of the information that are used by other procedures. Clearly, if different methods use different data portions, then different outcomes must be anticipated. So, a way to explain all voting anomalies is to codify these informational effects. With $N = 3$, this is accomplished by constructing an orthogonal basis for the data space (the six-dimensional profile space \mathbb{R}^6).

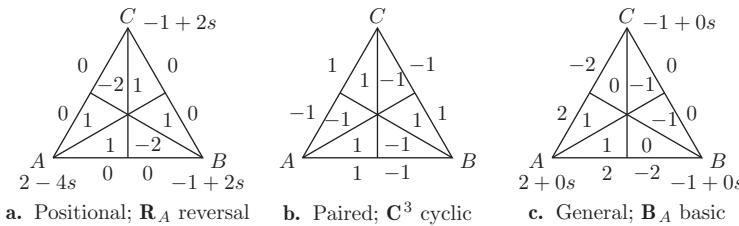
This decomposition identifies which data subspaces cause different methods to have different outcomes. To do so, basis vectors are defined that either affect *all* positional and pairwise rules in the same way, or have an impact on certain methods, but not others.

A first choice, which I call the *kernel* direction \mathbf{K}^3 , is where all rules have a compete tie. To define \mathbf{K}^3 , assign, for each of the six rankings, one agent. The α coefficient of a profile's kernel component, $\alpha\mathbf{K}^3$, satisfies $6\alpha = n$, where n is the number of agents. As such, expect the \mathbf{K}^3 coefficient to include fractions with denominator six. The space spanned by \mathbf{K}^3 is called the *kernel space* denoted by \mathbb{K}^3 .

Positional methods: The decomposition exploits the symmetries manifested by my geometric tallying method. As shown in Figure 7.3c, the difference between B 's plurality tally and B 's tally for any other positional is determined by entries in ranking regions 1 and 4 (the Figure 7.3c regions with “2nd”). Geometrically, these regions characterize a 180° rotation defining a pair consisting of a ranking and its reversal such as the indicated $\{A > B > C, C > B > A\}$. This reversal feature, of assigning one agent to each ranking in the pair, ensures that all paired comparisons are ties; e.g., if $X > Y$ is in the first ranking, then $Y > X$ is in the second, which leads to an $X \sim Y$ tie.

While reversal pairs ensure tied paired comparisons, something different happens with positional outcomes; e.g., for $\{A > B > C, C > B > A\}$, the \mathbf{w}_s outcome of $A:B:C$ has the $1:2s:1$ tally, where only the Borda count (with $s = \frac{1}{2}$) has a complete tie! Notice the bifurcation at $s = \frac{1}{2}$; the ranking for methods with $s < \frac{1}{2}$ reverses the ranking for methods with $s > \frac{1}{2}$! In particular, reversal components affect all non-Borda positional rankings and tallies, but they never affect pairwise rankings. As it turns out, *reversal components are completely responsible for all possible differences among positional method outcomes*.

To design examples, I use the reversal pairs as given above. For instance, to create Equation (7.1) with a $C > B > A$ paired comparison outcome, I assigned this $C > B > A$ ranking to two agents. If only reversal pairs are added, the paired comparison rankings remain unchanged. To obtain the desired positional differences, I used x pairs of $\{A > B > C, C > B > A\}$ and y of $\{A > C > B, B > C > A\}$. The resulting elementary algebra from the plurality and vote-for-two tallies determined the $x = 4$ and $y = 3$ values used to create Equation (7.1).

**Figure 7.4** A basis.

A formal representation of reversal terms is required to define the \mathbb{R}^6 basis. For technical reasons, these reversal components must be orthogonal to \mathbf{K}^3 , which requires the sum of the entries to equal zero. A natural choice is given in Figure 7.4a, which is \mathbf{R}_A^3 ; it assigns “1” to all rankings where A is top-ranked and in the corresponding reversal entry (where A is bottom-ranked). To be orthogonal to \mathbf{K}^3 , the -2 entries are added to the remaining reversal pair. Vectors \mathbf{R}_B^3 and \mathbf{R}_C^3 are similarly defined. According to Figure 7.4a, \mathbf{R}_X^3 vectors never affect paired comparisons. Don’t worry about the negative \mathbf{R}_X^3 components; they are handled by adding an appropriate multiple of \mathbf{K}^3 ; e.g., profile $\mathbf{R}_A^3 + 2\mathbf{K}^3$ has non-negative terms with the desired reversal properties. In this way, interpret the negative and positive entries of a \mathbf{R}_X vector as transferring $\alpha\mathbf{K}^3$ entries.

A straightforward computation shows that

$$\mathbf{R}_A^3 + \mathbf{R}_B^3 + \mathbf{R}_C^3 = 0, \quad (7.13)$$

so they span a two-dimensional *reversal* space denoted by \mathbb{RS}^3 . While vectors $\{\mathbf{R}_A^3, \mathbf{R}_B^3, \mathbf{R}_C^3\}$ are not orthogonal (it is easier to use \mathbf{R}_X^3 terms), the space \mathbb{RS}^3 is orthogonal to \mathbb{K}^3 .

Paired comparisons: A second triangular symmetry, the 120° difference, collects a ranking from every other ranking region. A first choice is

$$A \succ B \succ C, \quad B \succ C \succ A, \quad C \succ A \succ B, \quad (7.14)$$

and the remaining choice is

$$A \succ C \succ B, \quad C \succ B \succ A, \quad B \succ A \succ C. \quad (7.15)$$

In both sets, each alternative is in each position precisely once, so all positional outcomes for either set are complete ties; e.g., Equations (7.14) and (7.15) terms never affect positional rankings.

But paired comparisons of these terms define cycles. As earlier shown, the Equation (7.14) profile defines the

$$A \succ B, \quad B \succ C, \quad C \succ A$$

cycle with 2:1 tallies, while the Equation (7.15) profile defines the

$$B \succ A, \quad A \succ C, \quad C \succ B \quad (7.16)$$

cycle with the same 2:1 tallies.

To create examples, I use Equations (7.14) and (7.15). For instance, to construct a profile with the Equation (7.1) plurality, Borda, and vote-for-two rankings, but with a $A > C, C > B, B > A$ cycle, add z units of Equation (7.15) to Equation (7.1) to obtain

Number	Ranking	Number	Ranking	
$3 + z$	$A > C > B$	4	$A > B > C$	
3	$B > C > A$	$6 + z$	$C > B > A$	
z	$B > A > C$			

(7.17)

Equation (7.15) terms never affect positional rankings, so the Equation (7.17) positional rankings (and differences between tallies) remain the same as with Equation (7.1). However, Equation (7.17)'s pairwise tallies are $A:B$ with $(7+z):(9+2z)$, $B:C$ with $(7+z):(9+2z)$, and $A:C$ with $(7+2z):(9+z)$, which means that the rankings $C > B$ and $B > A$ hold for any $z \geq 0$. The desired $A > C$ outcome requires $z \geq 3$, so using $z = 3$ creates an example.

A formal representation of this cyclic term is to be orthogonal to \mathbb{K}^3 and \mathbb{RS}^3 . The \mathbf{C}^3 *cyclic* choice in Figure 7.4b places +1 in each of the Equation (7.14) regions and -1 in each of the Equation (7.15) regions. The one-dimensional \mathbb{R}^6 subspace spanned by \mathbf{C}^3 is the *cyclic space* denoted by \mathbb{CY}^3 . A computation proves that \mathbb{CY}^3 , \mathbb{RS}^3 and \mathbb{K}^3 are orthogonal to each other.

Basic terms: Identified are data components that allow nothing to happen (the one-dimensional space \mathbb{K}^3), that cause positional outcome differences without affecting paired comparison rankings (the two-dimensional space \mathbb{RS}^3), and that force differences in paired comparisons, even allowing cycles, but do not affect positional rankings (the one-dimensional cyclic space \mathbb{CY}^3 .) What remains is the two-dimensional subspace orthogonal to these spaces; I call it the *basic subspace*, denoted by \mathbb{BA}^3 , where “basic” refers to its fundamental status.

A formal definition of the basic vector in the A direction, denoted by \mathbf{B}_A^3 , is in Figure 7.4c. Here, +1 is in each region where A is top-positioned, 0 in each region where A is middle-positioned, and -1 in each region where A is bottom-positioned, and so it captures the sense of where voters with A bottom-ranked now have her top-ranked. Vectors $\mathbf{B}_B^3, \mathbf{B}_C^3$ are similarly defined. A direct computation shows that \mathbb{BA}^3 is orthogonal to \mathbb{K}^3 , \mathbb{RS}^3 , and \mathbb{CY}^3 . Because

$$\mathbf{B}_A^3 + \mathbf{B}_B^3 + \mathbf{B}_C^3 = \mathbf{0}, \quad (7.18)$$

these vectors span the two-dimensional \mathbb{BA}^3 basic subspace.

The properties of \mathbb{BA}^3 entries follow from Figure 7.4c; *nothing goes wrong!* Not only are all positional outcomes in agreement, but they must agree with the paired comparison outcomes! With a little algebra, it follows that the tally of any positional method (say, the plurality vote) of a profile in \mathbb{BA}^3 completely identifies the tallies for all other positional methods and all paired comparisons. Similarly, if the paired comparison tallies are known, the tallies for all positional methods can be determined. Although the outcome behavior is highly desired, the setting is rare because it belongs to the two-dimensional \mathbb{BA}^3 . The reality of how highly unlikely

it is to randomly select a Basic vector is captured by the fact that even a unanimity profile is *not* in \mathbb{BA}^3 .⁵

7.3.3.2 Summary

This decomposition can be used to completely describe all possible positional/paired comparison interactions whether they involve agreement or differences.

Theorem 7.6 (Saari, 1999) *For $N = 3$ alternatives, any profile can be uniquely decomposed into components in the one-dimensional (kernel) \mathbb{K}^3 , two-dimensional (reversal) \mathbb{RS}^3 , one-dimensional (cyclic) \mathbb{CY}^3 , and two-dimensional (basic) \mathbb{BA}^3 subspaces, which are mutually orthogonal to each other.*

The kernel component of a profile identifies the number of voters.

A profile's basic component has complete agreement among the rankings and tallies of all positional and paired comparison outcomes. This agreement allows the tally for one method to uniquely determine the tallies for all other methods.

The reversal component is completely responsible for any and all differences among positional methods; this includes differences in tallies and in rankings. The reversal component does not affect paired comparison rankings.

The cyclic component is completely responsible for all irregularities and differences among paired comparisons, which includes all cyclic outcomes. It does not affect differences in any positional methods tally of alternatives.

As a corollary, profile components in the \mathbb{CY}^3 and/or \mathbb{RS}^3 spaces cause all possible differences among positional and paired comparison outcomes. Theorem 7.6 creates examples illustrating that any admissible election anomaly reduces to simple algebra (Saari, 1999).

The dimensionality of these subspaces suggests what to expect. Consistency among all positional/paired comparison election outcomes, for instance, requires the profile to be near the lower-dimensional \mathbb{BA}^3 subspace, making it an unusual, rather than a typical setting. (This comment is supported by Theorems 7.3 and 7.4.) Namely, *anticipate at least some disagreement when tallying data with different methods*. This assertion is supported by personal experience where, in all elections with 10 or more voters (e.g., departments, committees, etc.) for which I have access to the original data (ballots with complete rankings of candidates), the outcomes would change with the tallying method. This includes a *society of social choice and behavior* election (Saari, 2001a).

This comment conflicts with claims about the surprising consistency of election outcomes in professional societies (e.g., Regenwetter *et al.*, 2007). If these assertions are correct, it would require (Theorem 7.6, related results, and the description in the opening paragraph of Section 7.4.5 indicating why voting outcomes can differ) the members of these organizations to have highly consistent preferences; e.g.,

⁵ If the unanimity profile for, say, $A > B > C$ were a basic vector, its plurality outcome ($A > B \sim C$), Borda outcome ($A > B > C$), and vote-for-two outcome ($A \sim B > C$) would agree, which is not true.

for $N = 3$, there would be no sizable “ $Ann > Barb > Connie; Connie > Barb > Ann$ ” division between the leading candidates of Ann and Connie. (Such an \mathbb{RS}^3 term would force different positional outcomes.) Thus, these profiles would belong to a very rare setting by being near the two-dimensional \mathbb{BA}^3 subspace.

A more reasonable explanation involves methodology. Access to complete sets of voter preferences from most of these societies was not available, so conclusions are *not* based on actual profiles. Instead, they are based upon “reconstructing” what *might* have been the actual data and profiles. This reconstruction method may not be accurate.⁶

7.3.4 Applications

The reason voting structures provide insight into what happens more generally is that aggregation rules typically include a step akin to voting, such as “select the item most often top-ranked” (e.g., the plurality vote). (Readers primarily interested in paired comparisons could jump to Section 7.4.) At this point voting structures provide tools and guidance about how to analyze the methods. To illustrate, two topics are selected: nonparametric statistics and Luce’s axioms for individual decisions.

7.3.4.1 Non-parametric Statistics

As with voting, different non-parametric statistical methods can yield different outcomes. In a standard hypothesis test, for instance, different approaches can have conflicting conclusions about the same data (e.g., see Bargagliotti & Saari, 2010). Some explanations follow almost immediately from voting results such as Theorems 7.5 and 7.6.

Haunsperger (1992) was among the first to explore the close theoretical connection between voting theory and non-parametric statistics. One of her conclusions, for instance, is that the Kruskal–Wallis method (Kruskal & Wallis, 1952) suffers many previously unsuspected inconsistencies and paradoxical behaviors. Then, she had a countering second conclusion, which is that of all possible non-parametric approaches, Kruskal–Wallis is, by far, the most consistent method!

The connection between voting and non-parametric statistics involves the way in which non-parametric rankings are found. To do so, the data are massaged into a form (i.e., data sets resembling profiles) where a desired positional voting or paired

⁶ It is clear from the decomposition that more information than normally available is required to even approximately reconstruct profiles. The Theorem 7.6 orthogonality conditions mean that a non-Borda positional outcome provides limited information about paired comparisons or even other positional results. With the $A > C > B$ plurality outcome and a 10:9:6 tally, for instance, it is impossible to determine whether the profile has 6 voters with $A > B > C$, 4 with $A > C > B$, 9 with $C > A > B$, 3 with $B > C > A$ and 3 with $B > A > C$, where all positional and paired rankings agree, or it is the Equation (7.17) profile with seven different positional outcomes (where each candidate is a “winner”) and a majority vote cycle.

comparison method can be used. This last step is what permits analyzing these approaches with results such as Theorems 7.5 and 7.6. (This approach of aggregation methods using positional methods is common, so results such as Theorems 7.5 and 7.6 are applicable in many other settings.)

To illustrate, consider hypothetical data from three companies, A , B , and C , producing a product where the “strength” variable is crucial

A	B	C	
4.2	5.1	3.2	
3.3	4.0	4.3	
3.7	3.6	3.8	

(7.19)

A class of non-parametric methods transforms this information into triplets with an entry from each column, such as $(4.2, 3.6, 4.3)$. Each Equation (7.19) column has three entries, so there are $3 \times 3 \times 3 = 27$ triplets. Each triplet is assigned a ranking according to the “strength” feature; e.g., $(4.2, 3.6, 4.3)$ is assigned the $C > A > B$ ranking. In this manner, associated with Equation (7.19) is the 27-agent profile

Number	Ranking	Number	Ranking	Number	Ranking
4	$A > B > C$	1	$A > C > B$	4	$C > A > B$
4	$C > B > A$	7	$B > C > A$	7	$B > A > C$

(7.20)

The Equation (7.20) information is aggregated by using a particular positional method; the tallies are used for other steps such as hypothesis testing, etc.

With Haunsperger’s connection between voting and statistics (part of her PhD thesis), she used Theorem 7.5 (and the above-described generalizations) to identify a shocking number of inconsistent paradoxical behaviors allowed by non-parametric methods. The Kruskal–Wallis test, for instance, is equivalent to the Borda Count, the Bhapkar V test (Bhapkar, 1961) is equivalent to the plurality vote, while the Bhapkar–Deshpandé L test (Bhapkar, 1968) uses what is equivalent to $(2, 1, 1, \dots, 1, 0)$, which is the Borda count for $N = 3$ (Haunsperger, 1992). Because of its identification with the Borda count, it immediately follows that the Kruskal–Wallis test has, by far, the smallest number of inconsistencies over rankings coming from the different sets of alternatives, the L test has limited consistencies, but only because it agrees with the Borda count for $N = 3$, while anything can happen with the V test.

Sharper voting theory results that now are available lead to a deeper understanding of non-parametric methods (Bargagliotti & Saari, 2010). With $N = 3$, for instance, the identification of non-parametric methods with voting rules permits using Theorem 7.6 to identify which data features cause conflict between, say, paired comparison and the Kruskal–Wallis outcomes. (This is totally due to \mathbb{CY}^3 terms.) It also provides a better understanding of different approaches. For instance, if reversal data of the $\{A > B > C, C > B > A\}$ type are significant, this supports using the V test (associated with the plurality vote) but with the accompanying cost of conclusions that potentially deviate from paired comparison

outcomes (because of \mathbb{RS}^3 and \mathbb{CY}^3 terms). For consistency (and any $N \geq 3$), the Kruskal–Wallis test is the optimal choice. An example in Bargagliotti and Saari (2010) shows how to use Theorem 7.6 structures to explain why different methods can reach different hypothesis-testing conclusions with the same data.

7.3.4.2 Luce's Axioms

An influential approach about human decision-making, which emphasized the structures of outcomes, was pioneered by Luce (1959). Luce left the source of these outcomes (i.e., the choice of the probability measure) and the structure of the space of inputs as variables to be determined. Because probability measures have features of voting rules, it is here that voting theory offers guidance.

Luce designed his approach to avoid those unrealistic requirements whereby all decision variables must be listed in advance. With his structure, alternatives can be added to, or subtracted from, the modeling without affecting major conclusions. As a consequence, his structure assigns an intrinsic probability to each element; e.g., A may be preferred twice as much as B independent of what other options are, or are not, available.

Luce's axiom is in terms of probability measures over certain spaces; i.e., $P_T(S)$ is the probability of S in T . For a pair, let $P(A, B)$ denote $P_{\{A,B\}}(A)$, which is the probability of selecting A from $\{A, B\}$.

Axiom 7.1 Luce's choice axiom (Luce, 1959) *For any $N \geq 2$; let $T = \{A_1, A_2, \dots, A_N\}$ be a set of N alternatives. A probability measure P_T satisfies the choice axiom if the following are true.*

- (1) *For every non-empty $S \subset T$, P_S is defined.*
- (2) *If $P(A_i, A_j) \neq 0, 1$ for all $A_i, A_j \in T$, then for $R \subset S \subset T$,*

$$P_T(R) = P_S(R)P_T(S). \quad (7.21)$$

- (3) *If $P(A_i, A_j) = 0$ for some $A_i, A_j \in T$, then for every $S \subset T$*

$$P_T(S) = P_{T-A_i}(S - A_i). \quad (7.22)$$

Axiom 7.1 imposes conditions on probability *values*; it is independent of the choice of a probability measure. As a comparison with voting, such an axiom would impose conditions on election tallies over different subsets of candidates without referral to what voting rule is being used. With the above, then, consequences in output space hold for all probability measures that satisfy the axiom. However, first, if $P(A, B) = 0$, then $P(B, A) = 1$, so Condition 3 refers to perfect discrimination; in voting it would be a unanimous victory. For what follows, assume imperfect discrimination, which means that $P(A_i, A_j) \neq 0, 1$ for any i, j .

As samples of what can be accomplished, following Luce (1959, pp. 68–74), let “ $P_T(A)$ denote the probability that A is judged to be the superior element in T according to some specified criterion.” Again, by not specifying what is a “superior element,” flexibility is introduced. Combining the axioms with standard rules of probability leads to the following (Luce has a more general representation):

Proposition 7.1 (Luce, 1959) *Assume the choice axioms hold. For $S \subset T$ and $A_i \in S$,*

$$P_S(A_i) = \frac{P_T(A_i)}{\sum_{A_j \in S} P_T(A_j)}. \quad (7.23)$$

For instance, if $T = \{A, B, C\}$, then $P(A, B) = \frac{P_T(A)}{P_T(A) + P_T(B)}$. A way to appreciate the power of this assertion is to describe what it would be with voting: If $P_T(X)$ represented the fraction of the total *plurality vote* won by X , then A 's vote against B in an n -person, three candidate majority vote would be $n \left(\frac{P_T(A)}{P_T(A) + P_T(B)} \right)$; a conclusion that normally is false.

Proposition 7.1 is powerful; in various forms it has been used even to explain errors in using the Nash strategy in game theory (e.g., see the quantal response approach developed in McKelvey & Palfrey, 1995 and an analysis of this approach in Jessie & Saari, 2016). This result endows each A_i with an intrinsic value where the likelihood of a particular alternative relative to others remains the same even with changes in the set of alternatives. A wishful goal in voting is for a profile's election rankings to be consistent over different subsets; Equation (7.23) is much stronger by ensuring consistency in rankings *and* in relative probability values.

The converse of Proposition 7.1, given in Equation (7.24), computes $P_T(A)$ in terms of paired comparisons. If this Equation (7.24) expression held for voting, it would assert that the plurality tally could be determined by majority votes over pairs; this assertion holds with basic terms but not in general.

Proposition 7.2 (Luce, 1959 p. 16) *If X, Y, Z represent the alternatives of $T = \{A, B, C\}$, then, under the assumption of imperfect discrimination,*

$$P_T(X) = \frac{P(X, Y)P(X, Z)}{P(X, Y) + P(X, Z) - P(X, Y)P(X, Z)} \quad (7.24)$$

and

$$P(T, C) = \frac{P(A, C)P(B, A)}{P(A, B)P(C, A) + P(B, A)P(A, C)}.$$

It is important to understand what happens with a specific aggregation rule, which is the choice of the probability measure. As these measures are related to positional voting rules, voting theory offers guidance about how to analyze the structures. (The connection between Axiom 7.1 and voting theory is in Saari, 2005; sample results follow.)

The measure Luce adopted is equivalent to the plurality vote; i.e., $P_T(A)$ is the sum of likelihoods of all rankings where A is top-positioned. This choice is as natural as adopting the plurality vote for elections. But, as now known, the plurality vote is an unreliable indicator of the voters' intent. Thus, while Axiom 7.1 prevents inconsistent outcomes, it is reasonable to anticipate that the weaknesses of the plurality vote are manifested elsewhere in satisfying Axiom 7.1. This is true; sample issues outlined here are:

- (1) The probability measure and Axiom 7.1 define the domain structure. Rather than a standard probability space, the structure is complex.
- (2) In constructing rankings, Luce discovered that outcomes can differ when starting with the best choice and working down from what happens when starting with the worst choice and working up. In particular, if $R(A \succ B \succ C)$ is the likelihood of ranking $A \succ B \succ C$, it is reasonable to expect that

$$R(A \succ B \succ C) = P_T(A)P(B, C), \quad (7.25)$$

which is the likelihood of A being the favorite times the likelihood that B is selected from the $\{B, C\}$ space. To compute the ranking by starting at the bottom, let $P_T^*(C)$ be the likelihood that C is the inferior choice; i.e., $P_T^*(C)$ is the sum of the probabilities of those rankings where C is bottom-ranked. With this measure, let $R^*(A \succ B \succ C) = P_T^*(C)P^*(B, A)$, or the likelihood C is an inferior choice times the likelihood B is judged inferior to A in $\{A, B\}$. With the obvious condition $P(A, B) = P^*(B, A)$, one might expect

$$R(A \succ B \succ C) = R^*(A \succ B \succ C). \quad (7.26)$$

But, in general, Equation (7.26) does *not* hold. Explanations come from voting theory. Intuition follows from the fact the plurality vote does not distinguish between a voter's second- and bottom-ranked candidate; Luce's probability choice inherits this weakness.

- (3) Can other probability measures eliminate some of these difficulties?

7.3.4.3 The Complex Domain Structure

To appreciate why the Luce domain structure is complex, start with standard rules of probability. With them, the five $N = 3$ variables $P_T(A), P_T(B), P(A, B), P(A, C), P(B, C)$ define the remaining terms. Each of the last three terms are related to the first two with an Equation (7.23) expression, so Luce's conditions define a smooth two-dimensional set. As the plurality and majority vote methods are smooth mappings, it follows from the implicit function theorem that the domain $\mathcal{D}_{\text{plurality}}^3$, which is the subset of probabilities that satisfy Luce's axiom with his plurality probability measure, is a two-dimensional manifold in Δ^5 , the five-dimensional probability space of three alternatives.

$\mathcal{D}_{\text{plurality}}^3$ is described in Saari (2005); the theme here is to show how Theorem 7.6 identifies what to investigate. For instance, had Luce been satisfied with consistency of rankings (rather than tallies) over the various subsets, then the Theorem 7.6 structures would have translated directly. Namely, $\mathcal{D}_{\text{plurality}}$ would be the basic subspace transferred into probability space \mathcal{BP}^3 where

$$\mathcal{BP}^3 = \left\{ \frac{a\mathbf{B}_A^3 + b\mathbf{B}_B^3 + c\mathbf{B}_C^3 + \alpha\mathbf{K}^3}{6\alpha} \mid \text{no component is negative; scalars } a, b, c, \alpha \right\}. \quad (7.27)$$

(Recall, the denominator 6α is the sum of the components of the numerator in Equation (7.27).)

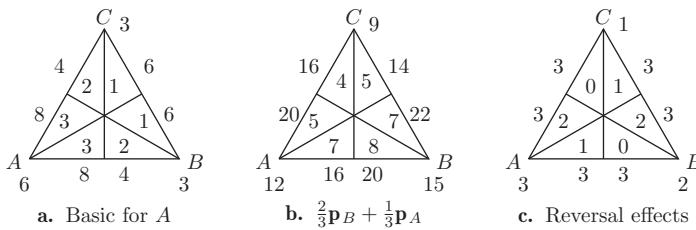


Figure 7.5 Domain conditions.

However \mathcal{BP}^3 is *not* the domain for Luce's structure. To see this with $\alpha = 1$, $\frac{\mathbf{B}_A^3 + \alpha \mathbf{K}^3}{6\alpha}$ defines the probability (with Equation (7.7) names) $\mathbf{p}_1 = (p_1, p_2, p_3, p_4, p_5, p_6) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, 0, 0, \frac{1}{6}) \in \mathcal{BP}^3$ that fails to satisfy Equation (7.23). Thus $\mathbf{p}_1 \notin \mathcal{D}_{\text{plurality}}$ which means that $\mathcal{D}_{\text{plurality}} \neq \mathcal{BP}^3$. In fact, *only* $\alpha = 2$ transfers $\frac{\mathbf{B}_A^3 + \alpha \mathbf{K}^3}{6\alpha}$ into the probability $\mathbf{p}_A = (\frac{1}{4}, \frac{1}{4}, \frac{1}{6}, \frac{1}{12}, \frac{1}{12}, \frac{1}{6}) \in \mathcal{D}_{\text{plurality}}^3$; \mathbf{p}_A , with the various tallies, are given in Figure 7.5a, where all integers have the denominator 12. Similarly, $\mathbf{p}_B = (\frac{1}{6}, \frac{1}{12}, \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{4}) = \frac{\mathbf{B}_B^3 + 2\mathbf{K}^3}{12} \in \mathcal{D}_{\text{plurality}}^3$, $\mathbf{p}_C = \frac{\mathbf{B}_C^3 + 2\mathbf{K}^3}{12} \in \mathcal{D}_{\text{plurality}}^3$.

Why is this so? Requiring consistent tallies (Equation (7.23)) forces $\mathcal{D}_{\text{plurality}}$ to have a complicated structure. In particular, $\mathcal{D}_{\text{plurality}}$ is not a linear space; it does not satisfy the standard property where a convex combination of $\mathcal{D}_{\text{plurality}}^3$ entries is in $\mathcal{D}_{\text{plurality}}^3$. The convex combination $\frac{1}{3}\mathbf{p}_A + \frac{2}{3}\mathbf{p}_B$, for instance, is *not* in $\mathcal{D}_{\text{plurality}}^3$! This probability (the numerators indicated in Figure 7.5b have the common denominator of 36) in \mathcal{BP}^3 satisfies Equation (7.23) for $\{A, B\}$, but not for $\{A, C\}$, nor $\{B, C\}$. For instance, according to Figure 7.5b, $P(A, B) = \frac{16}{36} = \frac{4}{9}$ while $P_T(A) = \frac{12}{36}$, $P_T(B) = \frac{15}{36}$, so $P(A, B) = \frac{4}{9}$ does equal $\frac{P_T(A)}{P_T(A)+P_T(B)} = \frac{12}{12+15}$. However $P(A, C) = \frac{20}{36} = \frac{5}{9} \neq \frac{P_T(A)}{P_T(A)+P_T(C)} = \frac{12}{12+9} = \frac{4}{7}$. Moreover, it is easy to show (with Equation (7.23)) that no $x \in (0, 1)$, α choices allow $\frac{1}{6\alpha}(x\mathbf{B}_A^3 + (1-x)\mathbf{B}_B^3 + \alpha\mathbf{K}^3) \in \mathcal{D}_{\text{plurality}}$. Instead, $\mathcal{D}_{\text{plurality}} \cap \mathcal{BP}^3 = \{\mathbf{p}_A, \mathbf{p}_B, \mathbf{p}_C\}$.

Nevertheless, to maintain agreement of rankings, $\mathcal{D}_{\text{plurality}}$ must be related to \mathcal{BP}^3 . To do so, reversal terms must be added to adjust the plurality tallies. However, reversal terms cannot affect differences in paired comparison tallies, so cyclic terms must also be added. The point to be made is that Theorem 7.6 provides guidelines of how to analyze the $\mathcal{D}_{\text{plurality}}$ structure (which is developed in Saari, 2005).

An unexpected complicating factor in satisfying the precise probability values in Axiom 7.1 is the seemingly innocuous kernel term $\alpha\mathbf{K}^3$. A way to avoid these difficulties is to subtract out the effects of the \mathbf{K}^3 component. That is, rather than $P_T(X)$, use how it differs from the average; e.g., $P_T(X) - \frac{1}{N}$. For pairs, this would be $P(X, Y) - \frac{1}{2}$, or $P(X, Y) - P(Y, X)$. (See the strongly transitive properties of Equation (7.31).) Doing so simplifies the domain; details are left to the interested reader.

7.3.4.4 Rankings: Top Down or Bottom Up

The $\mathcal{D}_{\text{plurality}}$ structure explains why, in general, $R(A \succ B \succ C) \neq R^*(A \succ B \succ C)$. The $P_T(A)$ ranking for $R(A \succ B \succ C)$ is equivalent to the plurality vote. However, voting for a voter's bottom-ranked candidate comes from the vote-for-two tally, so $P_T^*(C)$ outcomes reflect vote-for-two voting properties. Reversal terms, and only reversal terms, cause ranking changes between these methods. But $\mathcal{D}_{\text{plurality}}$ has reversal components, so differences must be expected. Indeed, with Figure 7.5c, $P_T(A) = \frac{1}{2}$, $P(B, C) = \frac{1}{2}$, so $P_T(A)P(B, C) = \frac{1}{4} \neq R(A \succ B \succ C) = \frac{1}{6}$. Even more, the reversal structure requires $P_T^*(C) = \frac{1}{6}$, $P^*(B, A) = P(A, B) = \frac{1}{2}$, so $R(A \succ B \succ C) = \frac{1}{6} \neq R^*(A \succ B \succ C) = (\frac{1}{6})(\frac{1}{2})$.

Although basic terms have the same tally for all positional methods, kernel elements do not. This means that $P_T(A)P(B, C)$ tallies are affected by \mathbb{K}^3 components, which means that even $B\mathcal{P}^3$ terms can experience difficulties. Using Figure 7.5a as an example, $R(A \succ B \succ C) = \frac{3}{12}$, which equals $P_T(A)P(B, C) = (\frac{6}{12})\frac{6}{12}$. However, $P_T^*(C) = \frac{3+2}{12}$ while $P^*(B, A) = P(A, B) = \frac{2}{3}$, which means that $R^*(A \succ B \succ C) = \frac{5}{18} \neq R(A \succ B \succ C)$. The culprit is the \mathbb{K}^3 component.

These computations identify why, in general, $R(A \succ B \succ C) \neq R^*(A \succ B \succ C)$; they are using different parts of the probability! To explain, $P_T(A) = p_1 + p_2$ while $P(B, C) = p_1 + p_5 + p_6$, so $P_T(A)P(B, C) = (p_1 + p_2)(p_1 + p_5 + p_6)$, while $P^*(C)P^*(B, A) = P^*(C)P(A, B) = (p_1 + p_6)(p_1 + p_2 + p_3)$. The p_5 value influences the $P_T(A)P(B, C)$ outcome, but not $P^*(C)P^*(B, A)$, and p_3 influences $P^*(C)P^*(B, A)$ values, but not $P_T(A)P(B, C)$. Similar to the Figure 7.3 geometric tally computations, if a variable appears in one computation but not another, differences must be expected. This explains this ranking difficulty.

Other probability measures: Results such as Theorem 7.6 make it clear that a probability measure patterned after the plurality vote must create complexities. There are many other choices; according to Theorem 7.6, a measure patterned after the Borda count could be expected to avoid many of the complexities identified by Luce. This is the case (Saari, 2005).

7.4 Paired Comparisons: An Analysis

Paired comparisons probably have been around since prehistoric combatants compared the size of a competing group versus their own. While widely used, many mysteries remain. For instance, how does one compare differences between Saaty's AHP and the Crawford and Williams geometric means approach? What causes cyclic outcomes, or all of those path-dependency issues that arise in psychology? While the list of concerns goes on, all reflect the reality that while examples can be designed to illustrate various effects, the missing step is a theoretical result identifying what causes *all possible paired comparison difficulties*.

It had been known since Condorcet (1785) that variants of Equations (7.6) and (7.14) cause cyclic outcomes. What was not known was whether other data structures could create paired comparison anomalies. This missing link limited

conclusions about how to avoid or mitigate paired comparison difficulties because it was not clear what other data structures could also cause difficulties.

Thanks to Theorem 7.6 this concern is completely resolved for $N = 3$; here, paired comparison rankings are affected by, and only by, \mathbb{BA}^3 and \mathbb{CY}^3 terms. The \mathbb{BA}^3 outcomes are surprisingly well-behaved, which means that *all possible $N = 3$ paired comparison difficulties and anomalies can be ascribed to \mathbb{CY}^3 terms*.

It remains to determine the source of everything that could possibly happen for $N \geq 3$. Answers are given in what follows by identifying the source of all possible paired comparison voting (and other aggregation method) anomalies. (These results were developed in Saari, 2000a, 2000b, 2014a.) To start, a general approach is given in terms of paired comparison outcomes. Of value for applications is that the source of this information, whether from voting, probability comparisons, etc., is immaterial. Then, a basis for voting rules is created.

7.4.1 Paired Comparison Outcomes⁷

For $N \geq 3$ alternatives, let $d_{i,j} \in \mathbb{R}$, $1 \leq i, j \leq N$, represent a comparison between A_i and A_j , where $d_{i,j} > 0$ indicates that $A_i \succ A_j$. The only constraint is that

$$d_{i,j} = -d_{j,i}, \quad (7.28)$$

which means that $d_{j,j} = 0$ for all j . In what follows, settings where summing $d_{j,k}$ terms is allowed are considered.

As examples, if $\tau(A_i, A_j)$ represents A_i 's tally in an $\{A_i, A_j\}$ majority vote election, then $d_{i,j} = \tau(A_i, A_j) - \tau(A_j, A_i)$ satisfies Equation (7.28); it measures how much better one candidate did than the other. Similarly, with the Section 7.3.4.2 probability notation, $d_{i,j} = P(A_i, A_j) - P(A_j, A_i)$ also works. As another example, if τ_i is the correlation of A_i with a base trait, such as IQ, then $d_{i,j} = \tau_i - \tau_j$ satisfies Equation (7.28). In other words, the power of examining consequences of $\{d_{i,j}\}$ terms is that results hold for *all* binary outcomes independent of from where they derive. In this way, conclusions from pairwise voting are intimately connected with assertions about probability, etc.

A way to make sense out of a data set $\{d_{i,j}\}$ is to assign a value, v_j , to A_j , $j = 1, \dots, N$, to designate A_j 's “strength” relative to the other alternatives. Other methods do this; e.g., with the $A \succ B \succ C$ plurality tally of 80:60:30, A is $80 - 30 = 50$ first-place votes better than C . If v_j values exist (and a goal is to define them), the $d_{i,j}$ data term could be replaced with

$$\bar{d}_{i,j} = v_i - v_j, \text{ which requires } \bar{d}_{j,i} = v_j - v_i = -\bar{d}_{i,j}, \bar{d}_{i,i} = v_i - v_i = 0. \quad (7.29)$$

Because of Equation (7.28), only $d_{i,j}$ values for $i < j$ need to be specified. So let the $\binom{N}{2} = \frac{N(N-1)}{2}$ dimensional space of binary outcomes be

$$\mathcal{BO}^N = \{\mathbf{d}^N = (d_{1,2}, d_{1,3}, \dots, d_{1,N}; d_{2,3}, \dots, d_{2,N}; \dots; d_{N-1,N}) \in \mathbb{R}^{\binom{N}{2}}\}. \quad (7.30)$$

⁷ The results in this subsection come from Saari (2014a), which has supporting details and added results.

The semicolons identify where the first subscript changes value.

If $d_{i,j}$ and $d_{j,k}$ have the same sign, then transitivity requires $d_{i,k}$ to have this sign. A stronger requirement, which I call *strongly transitive*, is that

$$d_{i,j} + d_{j,k} = d_{i,k} \text{ for all } i, j, k. \quad (7.31)$$

The pairwise tallies $A \succ B$ with 30:10, $C \succ B$ with 25:15, and $A \succ C$ with 25:15, for instance, satisfy strong transitivity because $d_{A,B} = \tau(A, B) - \tau(B, A) = 30 - 10 = 20$, $d_{B,C} = \tau(B, C) - \tau(C, B) = 15 - 25 = -10$, and $d_{A,C} = \tau(A, C) - \tau(C, A) = 10$, so

$$d_{A,B} + d_{B,C} = 20 - 10 = d_{A,C} = 10. \quad (7.32)$$

The strong transitivity equality (Equation (7.31)) provides clarity where transitivity is indecisive. With transitivity in Equation (7.32), for instance, the $A \succ B$ and $C \succ B$ rankings leave open the possibility of $A \succ C, A \sim C, C \succ A$; strong transitivity determines the choice.

Even more, should it be possible to realize the Equation (7.29) objective, then Equation (7.31) would mean that $\bar{d}_{i,j} + \bar{d}_{j,k} = (v_i - v_j) + (v_j - v_k)$ equals $\bar{d}_{i,k} = (v_i - v_k)$. This means that the Equation (7.29) goal of defining v_j values is intimately connected with strong transitivity.

7.4.1.1 Basis for \mathcal{BO}^N

According to Theorem 7.6, $N = 3$ pairwise outcomes are affected by basic and cyclic terms. The same holds for $N \geq 3$. As indicated next, \mathcal{BO}^N is orthogonally decomposed into the strongly transitive subspace, \mathcal{ST}^N , where nothing goes wrong, and the cyclic subspace, \mathcal{CY}^N , which causes all difficulties. (In Section 7.4.5, the basic \mathbb{BA}^N and cyclic \mathbb{CY}^N subspaces in profile space $\mathbb{R}^{N!}$ are mapped, respectively, onto \mathcal{ST}^N and \mathcal{CY}^N .)

The A_i basis vector for \mathcal{ST}^N is defined as follows:

Definition 7.3 The A_i basic vector, $\mathbf{Ba}_i^N \in \mathcal{BO}^N$, is where $d_{i,j} = 1$ for $j \neq i$, $j = 1, \dots, N$, and where $d_{k,j} = 0$ if $k, j \neq i$.

The $N = 4$ basic vectors are $\mathbf{Ba}_1^4 = (1, 1, 1; 0, 0; 0)$, where all $d_{1,j}$ terms equal one and all other terms equal zero, $\mathbf{Ba}_2^4 = (-1, 0, 0; 1, 1; 0)$ where the -1 reflects $d_{1,2} = -d_{2,1} = -1$, $\mathbf{Ba}_3^4 = (0, -1, 0 ; -1, 0; 1)$, and $\mathbf{Ba}_4^4 = (0, 0, -1; 0, -1; -1)$. Because $\sum_{j=1}^4 \mathbf{Ba}_j^4 = \mathbf{0}$, the vectors are not independent; they span a three-dimensional space. This is true in general.

Theorem 7.7 For $N \geq 3$, $(N - 1)$ -dimensional linear space \mathcal{ST}^N is given by the span of $\{\mathbf{Ba}_j^N\}_{j=1}^N$.

Later (Section 7.4.5) the \mathbb{BA}^3 basic space is extended to \mathbb{BA}^N . It will turn out that by using paired comparison tallies, \mathbb{BA}^N will be identified with \mathcal{ST}^N .

The proof of Theorem 7.7 is simple. Computations prove that each \mathbf{Ba}_j^N vector is strongly transitive and that any linear combination is strongly transitive. To prove that $\{\mathbf{Ba}_j^N\}_{j=1}^N$ is a basis for \mathcal{ST}^N , it suffices to show that \mathcal{ST}^N is a



Figure 7.6 Cyclic terms.

$(N - 1)$ -dimensional subspace of \mathcal{BO}^N ; this follows from the definition and associativity of addition. (See Saari, 2014a for details.)

The basis for the $\binom{N-1}{2}$ -dimensional cyclic linear space \mathcal{CY}^N , which contains all terms causing any kind of paired comparison difficulties, is appropriately associated with how terms are arranged around a circle. As indicated in Figure 7.6a, in any desired order, list the integers $1, 2, \dots, N$ in a clockwise manner around a circle's edge; this listing defines a permutation π . With Figure 7.6a, for instance, the listing is $\pi = \{2, 1, 4, 3\}$.

Definition 7.4 For $N \geq 3$ and permutation π , let \mathbf{C}_π^N be where $d_{i,j} = 1$ if j immediately follows i in π . (So, if j immediately precedes i , $d_{i,j} = -1$.) Otherwise, $d_{i,j} = 0$.

Illustrating with the Figure 7.6a choice of $\pi = (2, 1, 4, 3)$, $d_{2,1} = d_{1,4} = d_{4,3} = d_{3,2} = 1$; while $d_{2,4} = d_{1,3} = 0$. Thus $\mathbf{C}_{\{2,1,4,3\}}^4 = (-1, 0, 1; -1, 0; -1)$. The cyclic rankings associated with $\mathbf{C}_{\{2,1,4,3\}}^4$ are similarly determined; starting at the top of the Figure 7.6a circle and moving in a clockwise direction, 1 directly follows 2, so $d_{2,1} = 1$, which means that $A_2 > A_1$. Continuing around the circle, $A_1 > A_4, A_4 > A_3$, and $A_3 > A_2$ to complete the cyclic ranking. It is clear from the definition's use of a circle that each \mathbf{C}_π^N defines cyclic rankings.

Theorem 7.8 (Saari, 2014a) For $N \geq 3$, all cyclic vectors \mathbf{C}_π^N are orthogonal to the subspace \mathcal{ST}^N . The space spanned by all cyclic vectors equals \mathcal{CY}^N , which is the \mathcal{ST}^N normal subspace.

Theorem 7.8 finally identifies what causes *all possible paired comparison difficulties*; no matter what the troubles may be (e.g., cycles; where paired comparison rankings change by dropping or adding alternatives; path dependency issues), they are due to \mathcal{CY}^N terms and *only* \mathcal{CY}^N terms. This fact simplifies the analysis of paired comparison difficulties; e.g., to circumvent these problems (Section 7.5) design aggregation rules that neutralize the \mathbf{C}_π^N components.

According to Theorem 7.8, even paired comparisons that seem suitable for applications could have cyclic components! For instance, the pairwise elections of $A > B$ with 30:10, $C > B$ with 25:15, and $A > C$ with 25:15 are strongly transitive. However, changing the $A > B$ tally to 32:8 defines an outcome that is not strongly transitive, so it is the sum of strongly transitive and cyclic components. In other words, transitive outcomes that are not strongly transitive involve cyclic terms; the

cyclic components may be minor (relative to the strongly transitive terms), but the cyclic noise is there.

The proof of Theorem 7.8 is straightforward (see Saari, 2014a); e.g., the orthogonality condition is a direct computation. (The major part is to show that the span of $\{\mathbf{C}_\pi^N\}$ terms is a $\binom{N-1}{2}$ -dimensional linear subspace.) Namely, the effort to develop Theorem 7.8 was not the proof, but to discover the central role of the \mathbf{C}_π^N terms. There are, of course, more \mathbf{C}_π^N vectors than the $\binom{N-1}{2}$ dimension, so the cyclic terms have dependencies. To see this with $N = 3$, the cyclic space \mathcal{CY}^3 is one-dimensional, but there are six permutations $\{1, 2, 3\}, \{2, 3, 1\}, \{3, 1, 2\}$ and $\{1, 3, 2\}, \{3, 2, 1\}, \{2, 1, 3\}$. The first three permutations correspond to starting the $d_{i,j}$ computations at different positions of the Figure 7.6b wheel, so these \mathbf{C}_π^3 vectors equal $\mathbf{C}_{\{1,2,3\}}^3$. Those in the second set represent going in the counterclockwise direction about the wheel; for each of these π 's, $\mathbf{C}_\pi^3 = -\mathbf{C}_{\{1,2,3\}}^3$.

Many other bases can be created, even with triplets (Saari, 2014a, theorem 2), but the above choice is convenient for purposes of characterizing everything that can happen with paired comparisons.

7.4.2 Transitivity

An oft-asked question is whether the underlying transitivity for paired comparison problems can be restored. The answer is yes; a natural approach is to mimic “least squares” methods in statistics by projecting the data to eliminate noise. According to Theorem 7.8, the \mathbf{d}^N “noise” is its cyclic content, and so the goal is to find \mathbf{d}^N 's strongly transitive component. This projecting of \mathbf{d}^N into \mathcal{ST}^N makes sense; the \mathcal{ST}^N point closest to \mathbf{d}^N is \mathbf{d}^N 's \mathcal{ST}^N component.

While projecting \mathbf{d}^N into \mathcal{ST}^N removes \mathbf{d}^N 's cyclic terms, a much easier method uses the following Borda assignment rule,

Definition 7.5 (Saari, 2014a) For a given $\mathbf{d}^N, N \geq 3$, the *Borda assignment rule* (BAR) assigns to alternative A_i the value

$$\bar{b}_i = \sum_{j=1}^N d_{i,j}, \quad i = 1, \dots, N. \quad (7.33)$$

The $\{\bar{b}_i\}_{i=1}^N$ values define a transitive ranking of the alternatives.

An immediate consequence of Definition 7.5 and the $d_{i,j} = -d_{j,i}$ relationship is that $\sum_{j=1}^N \bar{b}_j = 0$. (To illustrate with $i = 4$, the $d_{4,5}$ term arises once when computing \bar{b}_4 , and a second time as $d_{5,4} = -d_{4,5}$ when computing \bar{b}_5 with $\sum_{j=1}^N \bar{b}_j$. These are the only two places where $d_{4,5}, d_{5,4}$ arise, and they cancel.)

Examples of BAR methods include the Borda count from voting and the Kruskal–Wallis approach (Kruskal & Wallis, 1952) from non-parametric statistics. An advantage of using BAR scores is that they drop all cyclic components; thus BAR is a projection. To see this with an example (from Saari, 2014a), let

$$\mathbf{d}^3 = \beta_1 \mathbf{B}\mathbf{a}_1^3 + \beta_2 \mathbf{B}\mathbf{a}_2^3 + c \mathbf{C}_{\{1,2,3\}}^3 = (\beta_1 - \beta_2 + c, \beta_1 - c; \beta_2 + c) \in \mathcal{BO}^3, \quad (7.34)$$

where $\mathbf{C}_{\{1,2,3\}}^3 = (1, -1; 1) \in \mathcal{CY}^3$. The BAR values are

$$\begin{aligned} \bar{b}_1 &= d_{1,2} + d_{1,3} = (\beta_1 - \beta_2 + c) + (\beta_1 - c) = 2\beta_1 - \beta_2, \\ \bar{b}_2 &= d_{2,1} + d_{2,3} = -d_{1,2} + d_{2,3} = (\beta_2 - \beta_1 - c) + (\beta_2 + c) = 2\beta_2 - \beta_1, \\ \bar{b}_3 &= d_{3,1} + d_{3,2} = -d_{1,3} - d_{2,3} = (-\beta_1 + c) + (-\beta_2 - c) = -\beta_1 - \beta_2. \end{aligned} \quad (7.35)$$

Of interest is how the c value, which is the coefficient for the cyclic term, drops from each \bar{b}_j . This cancellation happens in general because, for each \mathcal{C}_π^N and when computing \bar{b}_i , index i appears precisely twice in \mathcal{C}_π^N ; once with the j index that immediately follows i in π , and once with k that immediately precedes i ; i.e., $d_{i,j} = d_{k,i} = -d_{i,k} = 1$. As true with Equation (7.35), these two terms cancel in the BAR computation.

The general result (Saari, 2014a, theorem 3), which indicates the importance of the \mathcal{ST}^N component and β_j values in solving the stated objectives, such as Equation (7.29), follows.

Theorem 7.9 For $N \geq 3$, let $\{\mathbf{C}_{\pi_i}^N\}_{i=1}^{N-1}$ span \mathcal{CY}^N . Each \mathbf{d}^N has the unique representation (where $\beta_N = 0$)

$$\mathbf{d}^N = \sum_{i=1}^{N-1} \beta_i \mathbf{B}\mathbf{a}_i^N + \sum_{i=1}^{\binom{N-1}{2}} c_i \mathbf{C}_{\pi_i}^N. \quad (7.36)$$

For $1 \leq i \leq N$,

$$\bar{b}_i = (N-1)\beta_i - \sum_{k \neq i} \beta_k = \sum_{k \neq i} (\beta_i - \beta_k). \quad (7.37)$$

(By choice of the coordinate system, $\beta_N = 0$, so Equation (7.37) becomes $\bar{b}_N = -\sum_{i=1}^{N-1} \beta_i$.)

Similarly,

$$\bar{b}_i - \bar{b}_j = N(\beta_i - \beta_j), \quad (7.38)$$

and so $A_i \succ A_j$ if and only if $\beta_i > \beta_j$ if and only if $\bar{b}_i > \bar{b}_j$.

Selecting a \mathcal{ST}^N coordinate system is equivalent to choosing which $\beta_j = 0$. With $\beta_N = 0$ (so the coordinate system is $\{\mathbf{B}\mathbf{a}_j^N\}_{j=1}^{N-1}$), it follows (Equation (7.38)) that

$$\beta_j = \frac{1}{N}(\bar{b}_j - \bar{b}_N), \quad j = 1, \dots, N. \quad (7.39)$$

If v_j terms (Equation (7.29)) are defined, then, for constant e ,⁸

$$v_j = \frac{1}{N}\bar{b}_j - e = \beta_j - [e - \frac{1}{N}\bar{b}_N], \quad \bar{d}_{i,j} = \beta_i - \beta_j. \quad i, j = 1, \dots, N. \quad (7.40)$$

The projection of \mathbf{d}^N to \mathcal{ST}^N is $\bar{\mathbf{d}}^N = (\bar{d}_{1,2}, \dots, \bar{d}_{1,N}; \bar{d}_{2,3}, \dots; \bar{d}_{N-1,N})$.

⁸ This value is free to be selected; e.g., $e = 0$ would have $\sum_j v_j = 0$, while another choice could have the smallest v_j set equal to zero. Proofs are simple additions.

This theorem answers all of the objectives of this section, e.g., according to Equation (7.40), to find the sought-after v_j values, just subtract a fixed value from each β_j . To illustrate how Equations (7.39) and (7.40) define all relevant terms, consider $\mathbf{d}^4 = (1, 2, 3; -4, -5; 6)$, with BAR values $\bar{b}_1 = 1 + 2 + 3 = 6$; $\bar{b}_2 = -1 - 4 - 5 = -10$; $\bar{b}_3 = -2 + 4 + 6 = 8$; $\bar{b}_4 = -3 + 5 - 6 = -4$. The \mathbf{d}^4 strongly transitive component $\sum_{j=1}^3 \beta_j \mathbf{B} \mathbf{a}_j^4$ has $\beta_1 = \frac{1}{4}(\bar{b}_1 - \bar{b}_4) = \frac{1}{4}(6 - (-4)) = 2.5$, $\beta_2 = \frac{1}{4}(\bar{b}_2 - \bar{b}_4) = -1.5$, $\beta_3 = \frac{1}{4}(\bar{b}_3 - \bar{b}_4) = 3$. (Using β_j or the \bar{b}_j values shows that the \mathcal{ST}^4 part of \mathbf{d}^4 has the ranking $A_3 > A_1 > A_4 > A_2$.) The \mathcal{ST}^N component and β_j values answer the Equation (7.29) objectives giving $\bar{d}_{1,2} = \beta_1 - \beta_2 = 4$, $\bar{d}_{1,3} = \beta_1 - \beta_3 = -0.5$, $\bar{d}_{1,4} = \beta_1 - \beta_4 = 2.5$, $\bar{d}_{2,3} = -4.5$, $\bar{d}_{2,4} = \beta_2 = -1.5$, $\bar{d}_{3,4} = \beta_3 = 3$. Thus the projection of \mathbf{d}^4 is $\bar{\mathbf{d}}^4 = (4, -0.5, 2.5; -4.5, -1.5; 3)$. So, BAR, β_j , and $\bar{\mathbf{d}}^N$ values finally obtain goals such as extracting a data set's intended transitivity and attaching a value to each alternative.

7.4.3 Transitivity vs. Strong Transitivity

Theorem 7.9 provides new tools to investigate a range of issues. As an example, the following describes unexpected weaknesses of the standard transitivity condition.

Even if \mathbf{d}^N is transitive, its ranking need not agree with the strongly transitive ranking, which raises doubt as to whether the outcomes reflect the data's correct interpretation. To illustrate with Equation (7.34), the values $\beta_1 = 5$, $\beta_2 = 4$, and $c = -2$ define $\mathbf{d}^3 = (-1, 3; 2)$ with the $A_2 > A_1 > A_3$ transitive ranking. (As $d_{1,2} < 0$, $A_2 > A_1$; as $d_{1,3} > 0$, $A_1 > A_3$, and as $d_{2,3} > 0$, $A_2 > A_3$ to satisfy transitivity.) However, the BAR values $\bar{b}_1 = 2(5) - 4 = 6$, $\bar{b}_2 = 2(4) - 5 = 3$, $\bar{b}_3 = -5 - 4 = -9$ define the BAR ranking $A_1 > A_2 > A_3$ with $\bar{d}_{1,2} = 1$, $\bar{d}_{1,3} = 5$, $\bar{d}_{2,3} = 4$. The difference reflects how \mathbf{d}^3 's transitive ranking is distorted with cyclic terms; the BAR ranking, which agrees with the \mathcal{ST}^3 ranking of \mathbf{d}^3 (given by $\bar{\mathbf{d}}^3 = (1, 5; 4)$), removes the cyclic noise. (Also, the above $\mathbf{d}^4 = (1, 2, 3; -4, -5; 6)$ has the $A_1 > A_3 > A_4 > A_2$ ranking while $\bar{\mathbf{d}}^4$ has $A_3 > A_1 > A_4 > A_2$.)

BAR values finally accomplish the goal of removing the cyclic noise to achieve transitivity. However, perhaps other approaches can be designed that could claim to do the same. For instance, the Bhapkar V test (Bhapkar, 1961) (Section 7.3.4.1) attempts to do so. For other examples, mimic voting methods were developed to extract transitivity from cyclic pairwise outcomes. (One could, for example, select the transitive ranking with tallies that are nearest the \mathbf{d}^N values.) This suggests comparing BAR with all potential competitors.

Major rules either are linear mappings that map $\mathbf{d}^N \in \mathcal{BO}^N$ to numbers for each alternative, or have a major component that does so. As such, it suffices to consider linear mappings from \mathcal{BO}^N to \mathbb{R}^N . Next, A_j 's value should depend only on data involving A_j ; that is, the assigned value depends only on $d_{j,k}$ components of \mathbf{d}^N . Third, without cyclic noise terms (i.e., $\mathbf{d}^N \in \mathcal{ST}^N$), the ranking must agree

with that of the strongly transitive terms. Any method satisfying these conditions is called *pairwise faithful* (Saari, 2014a).

It is not difficult to create pairwise faithful rules. Of importance is that to avoid paradoxical behavior, the rule must not be influenced by cyclic terms. Expressed mathematically, \mathcal{CY}^N must be in the rule's kernel. As asserted next, BAR is the only such rule.

Theorem 7.10 (Saari, 2014a) *If a pairwise faithful map has \mathcal{CY}^N in its kernel, then the map is a positive multiple of BAR.*

Theorem 7.10 states that *if transitivity is desired, BAR is the only approach to use*. That BAR achieves the goals follows from Equations (7.38)–(7.40); what was not obvious is that BAR is, essentially, the only approach.

7.4.4 Seeking Perversities

Resembling the power of Luce's approach (e.g., Section 7.3.4.2), an advantage of developing consequences of $d_{i,j}$ values (without worrying about from where they come) is that results hold for a wide spectrum of rules ranging from voting to probability. This motivates seeking other $d_{i,j}$ properties, which then apply to a variety of aggregation rules. To illustrate, derive some paired comparison anomalies that plague all (additive) paired comparison methods.

The approach is simple. To explore what can happen with a particular linear method, it suffices (Theorems 7.8 and 7.9) to determine, separately, how the action affects strongly transitive and cyclic terms. Consequences are then discovered by combining terms. An issue (motivated by Theorem 7.5 where inconsistencies can arise by dropping alternatives and Proposition 7.1 where nothing goes wrong) is to examine what can happen with paired comparisons by removing or adding alternatives. The first step is to determine how this action affects strongly transitive and cyclic terms.

Dropping an alternative with a strongly transitive term leaves a strongly transitive term that is completely consistent with the original. This must be so because Equation (7.31) must hold for the remaining $d_{i,j}$ components. Illustrating with

$$\mathbf{d}_1^4 = 3\mathbf{B}\mathbf{a}_1^4 + 2\mathbf{B}\mathbf{a}_2^4 + \mathbf{B}\mathbf{a}_3^4 = (1, 2, 3; 1, 2; 1),$$

removing alternative A_4 drops the $d_{1,4} = 3$, $d_{2,4} = 2$, and $d_{3,4} = 1$ terms from \mathbf{d}_1^4 to define the strongly transitive

$$\mathbf{d}_1^3 = 2\mathbf{B}\mathbf{a}_1^3 + \mathbf{B}\mathbf{a}_2^3 = (1, 2; 1).$$

Proving that \mathbf{d}_1^3 is strongly transitive requires showing that $d_{1,2} + d_{2,3} = d_{1,3}$, but this is known because this expression must hold for \mathbf{d}_1^4 to be strongly transitive.

This projection property *fails* with cyclic terms. Illustrating with

$$\mathbf{d}_2^4 = \mathbf{C}_{\{1,2,3,4\}}^4 = (1, 0, -1; 1, 0; 1),$$

dropping A_4 defines $\mathbf{d}_2^3 = (1, 0; 1)$, which is not strongly transitive (as $d_{1,2} + d_{2,3} = 1 + 1 \neq d_{1,3} = 0$) nor cyclic (all $N = 3$ cyclic terms are multiples of $(1, -1; 1)$ and \mathbf{d}_2^3 does not have this form). According to Theorem 7.8, \mathbf{d}_2^3 is the sum of cyclic and strongly transitive terms. Using Equation (7.39) and \mathbf{d}_2^3 BAR values of $\bar{b}_1 = 1, \bar{b}_2 = 0, \bar{b}_3 = -1$, it follows that $\beta_1 = \frac{1}{3}(\bar{b}_1 - \bar{b}_3) = \frac{2}{3}$ and $\beta_2 = \frac{1}{3}(\bar{b}_2 - \bar{b}_3) = \frac{1}{3}$. Thus (with subtracting vectors),

$$\mathbf{d}_2^3 = \frac{1}{3}(2\mathbf{B}\mathbf{a}_1^3 + \mathbf{B}\mathbf{a}_2^3) + \frac{2}{3}\mathbf{C}_{\{1,2,3\}}^3 = \frac{1}{3}(1, 2; 1) + \frac{2}{3}(1, -1; 1) = \frac{1}{3}\mathbf{d}_1^3 + \frac{2}{3}\mathbf{C}_{\{1,2,3\}}^3.$$

Removing A_4 , then, converts the cyclic $\mathbf{C}_{\{1,2,3,4\}}^4$ into the cyclic $\mathbf{C}_{\{1,2,3\}}^3$, which is obtained by dropping 4 from $\pi = (1, 2, 3, 4)$, and a strongly transitive portion with $A_1 \succ A_2 \succ A_3$, which follow from π .⁹ This projection structure holds in general.

Theorem 7.11 (Saari, 2014a) *For $N \geq 4$, by removing alternative A_j , a strongly transitive term \mathbf{d}^N becomes a strongly transitive \mathbf{d}^{N-1} with entries that agree with the corresponding \mathbf{d}^N . However, a \mathbf{C}_π^N becomes the sum of strongly transitive and cyclic \mathcal{BO}^{N-1} terms. The cyclic term is $\mathbf{C}_{\pi^*}^{N-1}$ where π^* is the permutation defined by dropping index j from π .*

According to Theorem 7.11, if \mathbf{d}^N has cyclic components, dropping an alternative *must* alter the resulting data's strongly transitive portion. This is where the fun of finding unexpected consequences begins. Using $\mathbf{d}_1^j, \mathbf{d}_2^j$, for instance, suggests that, by dropping an alternative, the new strongly transitive component could *reverse the ranking!* Indeed, dropping A_4 with $\mathbf{d}_3^4 = \mathbf{d}_1^4 - 4\mathbf{d}_2^4$ and its strongly transitive $A_1 \succ A_2 \succ A_3 \succ A_4$ ranking leads to

$$\mathbf{d}_3^3 = \mathbf{d}_1^3 - 4\mathbf{d}_2^3 = -\frac{1}{3}\mathbf{d}_1^3 - \frac{8}{3}\mathbf{C}_{\{1,2,3\}}^3, \quad (7.41)$$

where the strongly transitive component now has the *reversed ranking* $A_3 \succ A_2 \succ A_1$. Similarly, with $\mathbf{d}_4^4 = \mathbf{d}_1^4 - 3\mathbf{d}_2^4$, dropping A_4 has $\mathbf{d}_4^3 = -2\mathbf{C}_{\{1,2,3\}}^3$, a cyclic vector with no strongly transitive terms! As these examples suggest, *anything* can happen!

Theorem 7.12 (Saari, 2014a) *For $N \geq 4$, select a transitive ranking, R_N , of the N alternatives, and another transitive ranking, R_{N-1} , of the $(N - 1)$ alternatives $\{A_1, \dots, A_{N-1}\}$ (so A_N is dropped). There exists $\mathbf{d}^N \in \mathcal{BO}^N$ where its strongly transitive portion has the R_N ranking, but by dropping A_N , the ranking defined by \mathbf{d}^{N-1} 's strongly transitive part is R_{N-1} .*

To appreciate the significance of Theorem 7.12, remember that a reliable decision method must depend, as much as possible, on the data's \mathcal{ST}^N portion. According to Theorem 7.12, *all such methods must admit peculiar behavior when adding or removing alternatives!* Thus, for instance, removing (or adding) an

⁹ All BAR values for \mathbf{C}_π^N are zero. Dropping A_j removes $d_{j,k}$ terms, so some sums now are non-zero, which creates the strongly transitive component. With $\mathbf{C}_{\{1,2,3,4\}}^4$, for instance, $\bar{b}_1 = d_{1,2} + d_{1,3} + d_{1,4} = 1 + 0 - 1$. Dropping A_4 changes this sum to the above $\bar{b}_1 = 1 + 0 = 1$.

alternative can cause the Kruskal–Wallis ranking to admit various other rankings! The Kruskal–Wallis approach has plenty of company because this assertion holds for *all additive methods*. According to Theorem 7.12, the sole source of these anomalies are \mathcal{CY}^N terms.

Why hasn't such behavior been previously identified? It has in special cases, but it was not recognized that this is a general phenomenon. To discover such behavior, one must suspect it can happen and know where to explore. However, theoretical information describing where to search for such outcomes was not previously available; Theorem 7.12 provides the answer – the problem is everywhere. (Also, this behavior has not been widely noted because, as indicated by the example, the data need a relatively strong cyclic component.)

By plaguing all methods, it is important to find ways to avoid these problems. Again, thanks to the theory, the answer is immediate and easy to implement: rather than continuing to use a given \mathbf{d}^N , which perpetuates the complexities caused by the cyclic noise, eliminate the problem by replacing \mathbf{d}^N with $\bar{\mathbf{d}}^N$. It is easy to compute $\bar{\mathbf{d}}^N$, with the reward of preserving behavior even when removing alternatives (Theorem 7.11).

Illustrating with $\mathbf{d}_3^4 = (-3, 2, 7; -3, 2; -3)$, the BAR values are $\bar{b}_1 = -3 + 2 + 7 = 6$, $\bar{b}_2 = 3 + (-3) + 2 = 2$, $\bar{b}_3 = -2 + 3 - 3 = -2$, $\bar{b}_4 = -7 - 1 + 3 = -6$. Thus (Equation (7.39)) $\beta_1 = \frac{1}{4}(\bar{b}_1 - \bar{b}_4) = \frac{1}{4}(6 - (-6)) = 3$, $\beta_2 = \frac{1}{4}(\bar{b}_2 - \bar{b}_4) = \frac{1}{4}(2 - (-6)) = 2$, $\beta_3 = \frac{1}{4}(\bar{b}_3 - \bar{b}_4) = \frac{1}{4}(-2 - (-6)) = 1$, $\beta_4 = 0$ with (Equation (7.40)) $\bar{\mathbf{d}}_3^4 = (1, 2, 3; 1, 2; 1)$. Dropping alternatives with $\bar{\mathbf{d}}_3^4$, rather than with \mathbf{d}_3^4 , preserves ranking and other information.

The above is but one of many features to be exploited; others are described in Saari (2014a) and still others have yet to be considered. As an example, interesting structures emerge by dropping each alternative from $\mathbf{C}_{\{1,2,3,4\}}^4$ to define four different data sets. These structures explain and show how to mitigate the Theorem 7.12 negative assertion.

7.4.5 Basis in Profile Space

Assertions now can be discovered for a variety of paired comparison methods, but what if a specific topic is of interest? Can these structures be transferred? They can by decomposing the domain into components equivalent to \mathcal{ST}^N and \mathcal{CY}^N . How to do so is illustrated with pairwise voting; the resulting structures are then used to discover other new conclusions.

7.4.5.1 Basic Terms

The paired comparison basis in profile space $\mathbb{R}^{N!}$ (discovered independently and earlier; e.g., Saari, 2000a, 2000b) follows the lead of Theorem 7.8. Starting with $N = 3$, the \mathbb{BA}^3 and \mathbb{CY}^3 spaces remain as described. The kernel for paired comparisons is the three-dimensional subspace spanned by \mathbb{K}^3 and \mathbb{RS}^3 as these terms cannot affect paired comparison rankings (Theorem 7.6).

For $N \geq 3$, the goal is to identify appropriate $\mathbb{R}^{N!}$ subspaces where paired comparison outcomes are mapped to \mathcal{ST}^N and to \mathcal{CY}^N . Everything else in $\mathbb{R}^{N!}$ corresponds to terms that change individual tallies, but cannot influence $d_{i,j}$ values. In what follows I do not develop consequences of the basic vectors, but they are defined first.

Definition 7.6 For $N \geq 3$ and alternative X , a \mathbf{B}_X^N basic vector is where $2(N - j) - (N - 1) = N - 2j + 1$ agents are assigned to each ranking where X is j th positioned, $j = 1, \dots, N$.

The Definition 7.6 values may seem to be strange, but, in fact, they resemble Borda count tallies, which assign $(N - j)$ points to a j th-ranked candidate. (The multiple of two avoids fractions.) While the $N = 4$ candidate Borda count uses 3, 2, 1, 0 points to tally ballots, the inflated 6, 4, 2, 0 values could also be used. To remove terms that do not affect pairwise rankings (e.g., kernel terms), subtract the average of assigned points from each value; the sum of remaining terms equals zero. So for $N = 4$, subtract 3 from each $2(4 - j)$ to obtain 3, 1, -1, -3; for $N \geq 3$, subtract $(N - 1)$ from each $2(N - j)$, which leads to the Definition 7.6 values. (With Definition 7.6, the Figure 7.4c values must be doubled.)

Four-alternative profiles cannot be geometrically represented in an equilateral triangle, but an equilateral pyramid (Figure 7.7a) works quite well. To convert the tetrahedron into a two-dimensional figure, choose a vertex (e.g., D at the top in Figure 7.7a) and cut along the three connecting edges. Folding out the faces leads to Figure 7.7b, which consists of four smaller triangles (thicker edges); each triangle has the six strict rankings of the associated triplet, where the remaining alternative is bottom-ranked. For instance, the small triangle with thick edges on the lower left has vertices A, C, D , so, as with Figure 7.2a, it represents their six rankings where the missing B is bottom-ranked.

In this way, each small triangle's ranking depends on how close it is to each vertex. The Figure 7.7b region slightly to the left of vertex A with an edge on the horizontal line, for instance, is closest to A , next closest to D , then B , and finally C , so it has the $A > D > B > C$ ranking. The \mathbf{B}_A^4 profile is geometrically represented in this figure.

The \mathbf{B}_A^4 paired comparison $\{X, Y\}$ tallies are computed in the following way: find the edge-connecting vertices X and Y . The perpendicular bisector of this line

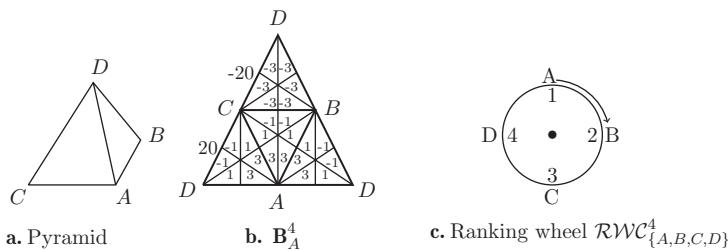


Figure 7.7 Basic and cyclic terms.

separates $X \succ Y$ from $Y \succ X$ preferences; so sum the values listed on each side of the line. Illustrating with $\{A, B\}$, the Figure 7.7b perpendicular bisector (of the line from A to B) connects C with the D in the lower right corner. All numbers below this line support A over B ; above the line, they support B over A . The tallies (posted next to the left edge of the folded-out pyramid) show that $A \succ B$ with a 20:–20 tally. (To illustrate with $\{A, D\}$, the perpendicular bisectors create a square with top vertices at C and D and bottom ones on the bottom edge. When Figure 7.7b is glued to create the pyramid, these two bottom points become attached.)

Computations show that any $\{A, X\}$ outcome has $A \succ X$ with a 20:–20 tally, but each $\{X, Z\}$ tally, $A_i \neq X, Z$, is a zero-to-zero tie. Thus, the \mathbf{B}_A^4 tallies define $40\mathbf{Ba}_1^4 = 40(1, 1, 1; 0, 0; 0)$. Thus, we have a desired basis that can be identified with \mathcal{ST}^N .¹⁰

Theorem 7.13 *The basic subspace, \mathbb{BA}^N , is the $(N - 1)$ -linear space spanned by $\{\mathbf{B}_{A_i}^N\}_{i=1}^N$. The majority vote $\mathbf{B}_{A_i}^N$ tallies for $\{A_i, Y\}$ are $A_i \succ Y$ with $M:−M$ where integer $M > 0$ depends on N . The $\{Z, Y\}$, $A_i \neq Y, Z$, tally is a 0:0 tie. These tallies identify $\mathbf{B}_{A_i}^N$, in a one-to-one manner, with $2M\mathbf{Ba}_i^N$. Thus, \mathbb{BA}^N is identified with \mathcal{ST}^N .*

All \mathbb{BA}^N vectors have strongly transitive pairwise outcomes. Of importance for applications is that \mathbb{BA}^N is isomorphic in \mathcal{ST}^N ; i.e., relationships in one space are paralleled with relationships in the other, so properties transfer back and forth. As an example, removing an alternative converts a \mathbb{BA}^N vector into a \mathbb{BA}^{N-1} vector, which alleviates concerns about whether this action introduces problems.

7.4.5.2 Cyclic Basis

Our goal is to understand what causes anomalies, so it is of particular interest to find the profile subspace that is identified with \mathcal{CY}^N . It should not be surprising that, as with \mathbf{C}_π^N , the *cyclic basis vectors* also are defined with a circle, which rotates to define a set of rankings that I call *ranking wheel configurations* (e.g., Saari, 2014b).

Start with a “ranking wheel,” which is a freely rotating wheel attached to a surface. With $N \geq 3$ alternatives, place, in an equally spaced manner around the wheel’s edge, the ranking numbers $1, 2, \dots, N$. (This is illustrated in Figure 7.7c with $N = 4$.) On the surface and next to a number, place the N names of the alternatives. (This can be done in any order; the order defines the listing $\pi = \{A_i, \dots, A_k\}$. The Figure 7.7c ordering is $\pi = \{A, B, C, D\}$.) The number next to a name defines the alternative’s ranking position; in Figure 7.7c, the ranking is $A \succ B \succ C \succ D$.

Next, rotate the ranking wheel to place “1” by the next name in π (in Figure 7.7c, this is B), and read off the new ranking. Continue until “1” has been placed next to each alternative in π precisely once. The resulting set of N rankings

¹⁰ I am leaving out some subtle points; e.g., the choice of \mathbf{B}_X^N must be such that removing an alternative defines a multiple of \mathbf{B}_X^{N-1} .

is called the *ranking wheel configuration* defined by π , and denoted by \mathcal{RWC}_π^N . With Figure 7.7c,

$$\mathcal{RWC}_{\{A,B,C,D\}}^4 = \{A \succ B \succ C \succ D, B \succ C \succ D \succ A, \\ C \succ D \succ A \succ B, D \succ A \succ B \succ C\}. \quad (7.42)$$

By construction, for each \mathcal{RWC}_π^N , each alternative is in first, second, ..., N th place precisely once, so for any positional method the outcome must be a complete tie. (Tversky's equation (7.3) has this same listing structure.) But for pairwise tallies, the outcome is a cycle where if A_j immediately precedes A_i in π , then A_j beats A_i by $(N - 1):1$. Illustrating with Equation (7.42), the majority vote tallies are $A \succ B, B \succ C, C \succ D, D \succ A$ each by 3:1.

It is easy to show that each \mathcal{RWC}_π^N is orthogonal to each $\mathbf{B}_{A_i}^N$.¹¹ To have the sum of the components equal to zero, the basis vector for cyclic effects is $\{\mathcal{RWC}_\pi^N, -\mathcal{RWC}_{-\pi}^N\}$ (where $-\pi$ is the reversal of π and $-\mathcal{RWC}_{-\pi}^N$ places a -1 for each $\mathcal{RWC}_{-\pi}^N$ ranking). Let \mathbb{CY}^N be the space spanned by $\{\mathcal{RWC}_\pi^N, -\mathcal{RWC}_{-\pi}^N\}$ over all possible listings π .

Theorem 7.14 *For $N \geq 3$, \mathbb{BA}^N is orthogonal to \mathbb{CY}^N . With pairwise voting, \mathbb{CY}^N is mapped onto \mathcal{CY}^N . Any profile (or profile component) in $\mathbb{R}^{N!}$ that is orthogonal to \mathbb{BA}^N and \mathbb{CY}^N has no impact on majority vote rankings.*

Theorem 7.14 finally identifies the source of *all possible pairwise voting difficulties*: they *are caused by, and only by, ranking wheel configurations* (and components). When this result was not known, it limited progress on analyzing pairwise voting outcomes. Thanks to Theorem 7.14, the analysis now can concentrate on what causes difficulties – \mathcal{RWC}_π^N terms. Some consequences are explored in Section 7.5.1.

As an example from the Tversky material in Section 7.1, consider the Equation (7.5) set of rankings $\mathbf{v} = \{A \succ B \succ C \succ D \succ E, E \succ D \succ C \succ B \succ A\}$. If each ranking has the same weight (e.g., same likelihood), then \mathbf{v} is orthogonal to \mathbb{BA}^5 and \mathbb{CY}^5 . (The first holds because when computing the scalar product with \mathbf{B}_X^5 , notice that alternative X is k th-ranked from the top in the first ranking and k th from the bottom in the other, so \mathbf{v} is orthogonal to each \mathbf{B}_X^5 . For the second, the first ranking defines $\mathcal{RWC}_{\{A,B,C,D,E\}}^5$ while the second defines the opposite, so the scalar produce equals zero.) As such (Theorem 7.14), paired comparisons of *these* two attribute vectors could not cause the Equation (7.2) outcome. However, with a richer attribute space, or with rankings from triples (e.g., the reversal components described in Theorem 7.6), the answer can change.

¹¹ In the N rankings of \mathcal{RWC}_π^N , A_i is in each position precisely once. When A_i is ranked k positions from the bottom, $\mathbf{B}_{A_i}^N$ assigns a value to that ranking that is the negative of what $\mathbf{B}_{A_i}^N$ assigns to a ranking where A_i is ranked k positions from the top, $k = 1, \dots, \frac{N}{2}$. Thus the $\mathbf{B}_{A_i}^N \cdot \mathcal{RWC}_\pi^N$ scalar product has a pairwise cancellation of terms.

7.5 Selective Consequences

Ways to use these paired comparison structures are suggested by describing new results. In Section 7.5.1, it is shown how knowing why paired comparison voting results occur permits resolving previous mysteries and significantly extending earlier conclusions. This includes Arrow's result (Section 7.2); resolutions of this 66-year-old mystery are described in Section 7.5.2. As indicated in Section 7.5.3, these structures can be used anywhere where paired comparisons arise; e.g., they even offer an alternative to the quantum thinking approach currently being used to analyze path-dependency issues.

My final example (Section 7.5.4) addresses the controversy about the competing scaled paired comparisons AHP and the geometric mean methods. (While I have described these results in lectures since 2010, this is the first published version.)

7.5.1 Voting: How Can These Results Be Used?

These voting rule structures allow several long-standing conclusions to be extended and difficulties to be resolved. A major one, starting with the debates between Borda and Condorcet in the 1780s, is to understand why the Borda ranking (favored by Borda) can differ from the pairwise majority vote rankings (favored by Condorcet, and, recently, by Maskin & Sen, 2016). A complete explanation follows.

A second illustration involves the large literature seeking ways to avoid majority vote difficulties, such as the Black single-peaked condition (Black, 1958). Why does this condition work? Can it be extended to handle supermajority methods, such as where a winner needs two-thirds of the vote, or, as in probability, support may require 95% agreement? By knowing the *only* cause of difficulties (Theorem 7.14), answers follow immediately.

7.5.1.1 Borda vs. Majority Vote Rankings

To compare majority voting and Borda, consider how a voter with $A > B > C > D$ preferences would vote over the six pairs.

Pair	A	B	C	D	
$\{A, B\}$	1	0	—	—	
$\{A, C\}$	1	—	0	—	
$\{A, D\}$	1	—	—	0	
$\{B, C\}$	—	1	0	—	
$\{B, D\}$	—	1	—	0	
$\{C, D\}$	—	—	1	0	
Total	3	2	1	0	

(7.43)

Each alternative's total tally is precisely the number of points assigned by the Borda count; this connection holds for all $N \geq 3$. In other words, the Borda count,

and only the Borda count, is the natural extension of pairwise voting. (For this reason, the majority vote over a pair can be treated as the $N = 2$ Borda outcome.)

Although simple, this Equation (7.43) relationship between the Borda count and pairwise outcomes explains why the Borda count is immune to many of the paradoxical outcomes identified by earlier theorems; Borda count outcomes must be connected to the paired comparison conclusions. By being a BAR (and this is where the “Borda” in BAR comes from), the Borda count is a projection that removes all cyclic terms.

Theorem 7.15 (Saari, 1999, 2000a, 2001b) *For $N \geq 3$, the pairwise majority vote and Borda count rankings agree if the profile is basic (strongly transitive). All differences between the pairwise vote and the Borda count are due to the pairwise vote tallies being affected by \mathbb{CY}^N terms. In any case, a Condorcet winner is always Borda ranked above a Condorcet loser, and, if the pairwise rankings are transitive, a Borda winner is always pairwise ranked above a Borda loser. Such consistency holds only for the Borda count; any other positional method can rank the candidates in an arbitrarily different manner than the majority vote rankings.*

In other words, the Borda count rankings (or BAR rankings) have all of the virtues of paired comparisons without suffering paradoxical difficulties. The reason the Borda count is spared problems is that it drops all cyclic noise terms generated by \mathbb{CY}^N components. And so the only reason majority vote outcomes can differ from the Borda ranking is that they are influenced by the cyclic noise, while the Borda count is not.

But, as true of all additive decision rules, the Borda count is subject to the Theorem 7.12 conclusion where rankings can change by dropping alternatives. The reason is that, by removing an alternative, a \mathbb{CY}^N term becomes the sum of \mathbb{BA}^{N-1} and \mathbb{CY}^{N-1} terms. Illustrating with $\mathcal{RWC}_{\{A,B,C,D\}}^4$

$$A \succ B \succ C \succ D, B \succ C \succ D \succ A, C \succ D \succ A \succ B, D \succ A \succ B \succ C,$$

dropping D leads to $A \succ B \succ C, B \succ C \succ A, C \succ A \succ B, A \succ B \succ C$, which is $\mathcal{RWC}_{\{A,B,C\}}^3$ plus an extra $A \succ B \succ C$. And so, as with Equation (7.41), the Borda ranking could change by removing an alternative if the profile has a sufficiently strong \mathbb{CY}^N component. Of interest, this is the only source of all inconsistencies allowed by the Borda count (and BAR methods)! (This is formally asserted in Theorem 7.16.) Other methods can be affected by this Condorcet behavior as well as other profile terms, such as reversal effects.

Theorem 7.16 *For $N \geq 3$, removing an alternative can change the Borda count ranking; this behavior is strictly due to the profile’s \mathbb{CY}^N terms. But by removing the cyclic noise from a profile, the Borda outcomes for all subsets of candidates agree with that of the N -candidates. Even with the cyclic noise in a profile, for any integer k , $2 \leq k < N$, summing each candidate’s k -alternative Borda tally equals a fixed multiple of each candidate’s N -candidate Borda tally.*

7.5.1.2 Avoiding Cycles

(What follows is from Saari, 2014b, which has other results and supporting arguments.) The social choice community is interested in ways to avoid cycles. A well-known method is Black's single-peaked condition (Black, 1958), which, for $N = 3$, requires that some candidate never is bottom-ranked. If the lucky alternative is C , then no voter has the $A > B > C$ or $B > A > C$ ranking; illustrating with Figure 7.2a, no voter can have preferences in regions 1 or 6. If Black's condition is satisfied, then, no matter how many voters and how their preferences are allocated, cycles cannot occur.

Black's condition, which prevents cycles by restricting the admissible preferences, does not prevent supermajority voting cycles. This weakness presented a previously open challenge to find appropriate conditions that would work with these methods.

Definition 7.7 With $n \geq 2$ voters, a (q, n) -rule is where, in a paired comparison, a winning alternative receives at least the quota, q , votes, where $q > \frac{n}{2}$. The majority vote is where q is the first integer greater than $\frac{n}{2}$; a two-thirds supermajority vote is where q is the first integer greater than, or equal to, $\frac{2n}{3}$; unanimity is where $q = n$.

In the US Senate, for instance, legislative bills may require 60 votes, so, with 100 senators, this is a $(60, 100)$ -rule vote. The problem is that non-unanimity (q, n) -rules can have cycles! In probability, for instance, it is possible to have cycles even if the probability of selecting X over Y must be at least 0.97. What rules prevent cycles?

Ward's condition: What simplifies analyzing potential rules is that (Theorem 7.14) *all possible paired comparison problems are caused by \mathcal{RWC}_π^N terms!* Thus, these issues reduce to limiting the damage caused by ranking wheel configurations. To illustrate with Black's condition, there are two \mathcal{RWC}_π^3 choices; one has a positive entry in Figure 7.2a regions 1, 3, 5, while the other has a positive entry in Figure 7.2a regions 2, 4, 6. If A never is bottom-ranked, the zeros in regions 4 and 5 prevent both \mathcal{RWC}_π^3 from being in the profile. Similarly, if B never is bottom-ranked, there are zeros in regions 2 and 3, and if C never is bottom-ranked, there are zeros in regions 1 and 6; both cases prevent a profile from having a \mathcal{RWC}_π^3 ; this is what prevents cycles from occurring.

More is possible: \mathcal{RWC}_π^3 terms can be excluded by requiring that nobody has preferences in at least one odd-numbered and one even-numbered region. Ward (1965) discovered these conditions in a different manner (because Theorem 7.14 was not known); he found that if nobody ranks some alternative in first place, *or* nobody ranks some candidate in second place, *or* nobody ranks some candidate in last place, this prevents cycles. In terms of Theorem 7.14, Ward's condition works by preventing the profile from including a ranking wheel configuration. And so we not only know (Saari, 2014b) why Ward's conditions work, but it now follows (Theorem 7.14) that his conditions are the best possible $N = 3$ constraints!

Nakamura number: Nakamura (1975, 1978) showed that limiting the number of alternatives can prevent (q, n) -rule cycles. With Theorem 7.14, this question becomes to find the smallest N where \mathcal{RWC}_π^N can generate a (q, n) -rule cycle; for a smaller number of alternatives, a cycle cannot arise. With the $(80, 100)$ -rule, for instance, can there be a six-alternative cycle?

A six-alternative cycle is created by \mathcal{RWC}_π^6 for $\pi = \{A, B, C, D, E, F\}$, which defines the $A > B, B > C, C > D, D > E, E > F, F > A$ cycle each with a 5:1 tally. Each pair's winner needs at least 80 votes, so $\frac{80}{5} = 16$ copies are needed. Each \mathcal{RWC}_π^6 has six voters, so these 16 copies require $6 \times 16 = 96$ voters. With 100 voters, such a cycle can be created.

The general argument uses the $(N - 1) : 1$ tallies from \mathcal{RWC}_π^N to find the number of copies, $\frac{q}{N-1}$, of \mathcal{RWC}_π^N needed to create a (q, n) -rule cycle. Next, determine whether there are enough voters; e.g., is $N(\frac{q}{N-1}) \leq n$? If so, a cycle can be created; if not, it cannot.

Illustrating the $(80, 100)$ -rule with $N = 4$, a \mathcal{RWC}_π^4 defines a 3:1 cycle. Thus $\frac{80}{3} > 26$ copies of \mathcal{RWC}_π^4 are needed for the $(80, 100)$ -rule cycle. These copies require more than $4(26) = 104$ voters, but only 100 are available. And so, with four alternatives, an $(80, 100)$ -rule cycle is impossible.

Collecting terms, and handling non-integer settings such as $\frac{80}{3}$ leads to a new, transparent, and simpler proof of the conclusion:

Theorem 7.17 (Nakamura, 1975, 1978) *For a (q, n) -rule, let $v(q, n) = \lceil \frac{n}{n-q} \rceil$ (which, if $\frac{n}{n-q}$ is not an integer, round up to the first larger integer). A $v(q, n)$ -alternative (q, n) -cycle can be created. However, if $N < v(q, n)$, then a (q, n) -rule cycle can never occur.*

As $v(80, 100) = \frac{100}{100-80} = 5$, a five-alternative $(80, 100)$ -rule cycle can be created, but four alternative cycles do not exist.

Supermajority profile restrictions: As a final illustration, while the Black, Ward, and natural extensions prevent majority vote cycles, nothing substantive had been discovered for supermajority (i.e., (q, n) -rule) settings. What prevents, for example, a $(60, 100)$ rule cycle? Thanks to Theorem 7.14, these concerns now are easy to resolve: prevent certain kinds of ranking wheel configurations from being in the profile. This leads to the following:

Theorem 7.18 (Saari, 2014b, theorem 4) *For $N \geq 3$ alternatives, a necessary and sufficient condition for a set of rankings to avoid a (q, n) -rule cycle, independent of how voters are assigned to preference rankings, is that, when restricted to each subset of $v(q, n)$ alternatives, at least one ranking is missing from each possible $\mathcal{RWC}_\pi^{v(q,n)}$.*

Returning to the US Senate $(60, 100)$ -rule, because $v(60, 100) = \lceil \frac{100}{100-60} \rceil = 3$, a necessary and sufficient condition that a set of rankings avoids a $(60, 100)$ -rule cycle (independent of how the voters assume rankings) is that when the set is restricted to a triplet, one of the Ward conditions apply. (Ward conditions, given

above, are equivalent to prohibiting a profile from including a \mathcal{RWC}_π^3 , so the assertion follows from Theorem 7.18.) For the (80, 100)-rule (where $v(80, 100) = 5$), the necessary and sufficient condition is that, when restricting the rankings to any quintuplet, at least one ranking is missing from each possible \mathcal{RWC}_π^5 .

7.5.2 Resolving Arrow's Theorem

New conclusions can be discovered by using Theorem 7.14 and \mathcal{RWC}_π^N ; the same approach holds to find ways to circumvent Arrow's theorem (Section 7.2).

Theorem 7.19 (Saari, 2000a) *For $N \geq 3, n \geq 2$, if a profile has no \mathbb{CY}^N components, then the Borda count, and majority vote pairwise rankings satisfy IIA, weak Pareto, and the non-dictatorial conditions while having complete, transitive outcomes. For Arrow's negative conclusion to hold, a profile must have a strong \mathbb{CY}^N component.*

So, to sidestep Arrow's negative conclusions, ways need to be found to minimize the influence of a profile's \mathcal{RWC}_π^N terms. But IIA (or the reductionist approach) forces attention on what happens with each pair, independent of other choices, so IIA ignores the influence of ranking wheel configurations. A goal, then, is to resurrect this information, and, if possible, to do so in a manner that maintains the spirit of Arrow's conditions.

A first approach recognizes the significant difference between having $A > B$ in $C > D > A > B$ and in $A > C > D > B$. In the second ranking, the intensity of the $A > B$ choice is of obvious importance; this leads to the following:

Definition 7.8 In a complete transitive ranking with $N \geq 3$ alternatives, the *intensity version of IIA* (IIIA) of an $X > Y$ ranking, denoted by $(X > Y, k)$, is where X is ranked k alternatives above Y .

With the above, Arrow's IIA treats $A > B$ the same in both rankings. In contrast, the IIIA information is $(A > B, 1)$ in the first ranking and $(A > B, 3)$ in the second. By replacing Arrow's IIA with IIIA, a positive assertion replaces Arrow's negative conclusion.

Theorem 7.20 (Saari, 1995a) *For $N \geq 3$ alternatives and $n \geq 2$ agents, if each agent has a complete transitive ranking over the alternatives (with no restrictions on the choice), then the Borda count satisfies IIIA and weak Pareto (i.e., if everyone prefers $X > Y$, that is the societal outcome), and the outcome is complete and transitive. (Indeed, the tallies are strongly transitive.) The Borda count is the only positional method with this property.*

To appreciate how IIIA handles ranking wheel configurations, consider $A > B > C > D, B > C > D > A, C > D > A > B, D > A > B > C$. The admissible IIA information has three $A > B$ rankings and one $B > A$, which indicates an $A > B$ outcome. However, with IIIA, the information has three $(A > B, 1)$ and one $(B > A, 3)$, where summing intensities leads to the tie expected from a \mathcal{RWC}_π^4 .

The Borda count arises in Theorem 7.20 because, for each ballot, the difference of Borda tallies between candidates is the intensity value.

A second approach retains Arrow's IIA, but treats the paired comparison information as the penultimate, rather than final step. This information, for instance, could be summed, such as with the Borda count where X 's tally is the sum of her tallies over all majority votes. This summation removes a profile's troubling \mathcal{RWC}_π^N components (Theorem 7.16).

Theorem 7.21 *If, in Arrow's theorem, information from IIA is treated as a penultimate step in finding the societal ranking, where the actual societal ranking is a linear combination of IIA conclusions, then the Borda count satisfies all of the conditions.*

As both theorems demonstrate, an efficient, simple way to avoid paired comparison difficulties is to use the Borda count. These conclusions can be treated as consequences of Theorem 7.14, which identifies the source of all paired comparison difficulties – ranking wheel configurations.

7.5.3 Path-dependency and Quantum Thinking

To discover other settings where these results can be applied, try anywhere where paired comparisons arise. To illustrate, Wang *et al.* (2014) addressed path-dependency problems coming from Gallup Poll data where events were evaluated in different orders. As a 1997 example concerning the perceived honesty of President Clinton and Vice President Gore, the comparison question was asked either about the honesty of Clinton or Gore, or in the other order; the order affected the outcome.

Wang *et al.* (2014) analyzed this dependency concern by using their favorite tool of “quantum thinking” to derive what they called the quantum question equalities (QQ equalities) indicated below. Because path-dependency concerns involve paired comparisons, it is reasonable to examine the data with \mathcal{BO}^N structures (Sections 7.4.1–7.4.4) to seek an alternative, simpler argument for the QQ equations. The following description, which comes from Saari (2016), shows that the QQ expressions are equivalent to the data not being subjected to serious cyclic effects. (Check Saari (2016) for details, proofs, and other conclusions.)

In the Clinton–Gore setting, the four alternatives are

- X_1 is a positive opinion of Clinton, X_2 is a positive opinion of Gore,
- X_3 is a negative opinion of Clinton, X_4 is a negative opinion of Gore.

Let $s_{i,j}$ be the fraction of people with X_i as the first outcome and X_j as the second; e.g., $s_{1,4} = 0.0447$ is the fraction of people viewing Clinton as honest (X_1) and Gore as dishonest (X_4) when asked in that order. In contrast, $s_{4,1} = 0.0255$ holds when asked about Gore first and Clinton second. Order effects are captured by defining $d_{i,j} = s_{i,j} - s_{j,i}$. Thus $d_{1,4} = 0.0447 - 0.0255 = 0.0192$.

Order effects fail to exist if $\mathbf{d}^4 = \mathbf{0}$, which means that how \mathbf{d}^4 deviates from zero identifies how the order matters. The Gallup information (Wang *et al.*, 2014) defines the context vector

$$\mathbf{d}_G^4 = (-0.0726, 0, 0.0192; 0.0224, 0; 0.0786), \quad (7.44)$$

where zeros represent unavailable, but presumably accurate values about the same person; e.g., $d_{1,3}$ is where Clinton is viewed as honest and then dishonest. The \mathbf{d}_G^4 BAR values are $\bar{b}_1 = -0.0534$, $\bar{b}_2 = 0.0950$, $\bar{b}_3 = 0.0562$, $\bar{b}_4 = -0.0978$ with the $X_2 \succ X_3 \succ X_1 \succ X_4$ ranking. Of interest is how the $X_2 \succ X_4$ inequality is separated by the $X_3 \succ X_1$ negative assessment of Clinton; this $(X_2 \succ X_4, 3)$ intensity ranking (Definition 7.8) captures the strong, positive opinion of Gore relative to that of Clinton.

In addition to $d_{1,3} = d_{2,4} = 0$, another constraint is that the sum of probabilities where Clinton is listed first must equal unity, $[s_{1,2} + s_{1,4}] + [s_{3,2} + s_{3,4}] = 1$, similarly for Gore, $[s_{2,1} + s_{2,3}] + [s_{4,1} + s_{4,3}] = 1$. This defines the equality

$$d_{1,2} + d_{1,4} - d_{2,3} + d_{3,4} = 0. \quad (7.45)$$

Any other independent equality would split Equation (7.45) into two equalities. A natural choice asserts that the relevant variables have no cyclic effects. This is

$$0 = \langle \mathbf{d}^4, \mathbf{C}_{(1,2,3,4)}^4 \rangle = d_{1,2} + d_{2,3} + d_{3,4} - d_{1,4}. \quad (7.46)$$

(For \mathbf{d}_G^4 , the right-hand side of Equation (7.46) equals 0.0092.) Solving Equations (7.45) and 7.46 leads to $d_{1,2} + d_{3,4} = 0$, which requires YY path effects to be countered by NN effects, and $d_{1,4} = d_{2,3}$, which means that NY and YN values agree. *These are the QQ equations proposed in Wang *et al.* (2014)!* Thus this transitivity equation (Equation (7.46)), which mandates an absence of cyclic effects in the relevant data, is equivalent to the QQ expressions!

Theorem 7.22 (Saari, 2016, theorem 10) *If $\mathbf{C}_{(1,2,3,4)}^4$ is orthogonal to \mathbf{d}^4 (Equation (7.46)), then the QQ equalities hold and \mathbf{d}^4 has the idealized QQ vector form*

$$\tilde{\mathbf{d}} = (-\alpha, 0, \beta; \beta, 0; \alpha). \quad (7.47)$$

Conversely, if QQ holds, then so does Equation (7.47); its decomposition is

$$\tilde{\mathbf{d}} = \frac{\beta}{2} \mathbf{B}_1^4 + \frac{\alpha + \beta}{2} \mathbf{B}_2^4 + \frac{\alpha}{2} \mathbf{B}_3^4 - \frac{\alpha}{2} \mathbf{C}_{(1,2,4,3)}^4 + \frac{\beta}{2} \mathbf{C}_{(1,4,2,3)}^4, \quad (7.48)$$

so $\tilde{\mathbf{d}}$ is orthogonal to $\mathbf{C}_{(1,2,3,4)}^4$. The Equation (7.47) BAR values are $\bar{b}_1 = \beta - \alpha$, $\bar{b}_2 = \beta + \alpha$, $\bar{b}_3 = \alpha - \beta$, $\bar{b}_4 = -\alpha - \beta$.

By using the structures described here, the main result in Wang *et al.* (2014) admits a simple, somewhat standard but very different interpretation (without reference to quantum thinking); it is that *the relevant variables are not seriously affected by cyclic noise terms!*

7.5.4 Consistency in Ratio-scaled Comparisons

The above structures are not appropriate should paired comparisons be combined in a non-additive manner, but they remain very useful if the additive operations can be identified with a group isomorphism. To indicate how to do so, a multiplicative approach is described.

In paired comparisons, let $a_{i,j}$ be the multiple of how much better A_i is than A_j . Clearly

$$a_{i,j} > 0, \quad a_{j,i} = \frac{1}{a_{i,j}}, \quad a_{i,i} = 1. \quad (7.49)$$

With Equation (7.49), only $a_{i,j}$ terms with $i < j$ need to be specified, so the space of *scaled comparisons* can be given by

$$\mathcal{SC}^N = \{\mathbf{a}^N = (a_{1,2}, \dots, a_{1,N}; a_{2,3}, \dots, a_{2,N}; \dots, a_{N-1,N}) \mid a_{i,j} > 0\}.$$

A way to analyze given data \mathbf{a}^N is to assign weights to each alternative; w_i to A_i , where larger weights identify stronger alternatives. The intent is to have

$$a_{i,j} = \frac{w_i}{w_j}. \quad (7.50)$$

If $\mathbf{w}^N = (w_1, \dots, w_N)$ exists, then $a_{i,j}a_{j,k} = \left(\frac{w_i}{w_j}\right)\left(\frac{w_j}{w_k}\right) = \frac{w_i}{w_k} = a_{i,k}$. Thus, a *consistency condition*, related to the existence of \mathbf{w}^N , is

$$a_{i,j}a_{j,k} = a_{i,k} \quad \text{for all } i, j, k. \quad (7.51)$$

Let \mathcal{CN}^N be the $(N - 1)$ -dimensional “consistency manifold” of points $\mathbf{a}^N \in \mathcal{SC}^N$ that satisfy Equation (7.51). A necessary condition to define \mathbf{w}^N for $\mathbf{a}^N \in \mathcal{CN}^N$.

The associated set of issues is clear. How can \mathbf{w}^N be determined, what does it mean, and how can it be defined for $\mathbf{a}^N \notin \mathcal{CN}^N$? Saaty’s AHP method defines \mathbf{w}^N in terms of the eigenvectors for the $N \times N$ matrix $\mathcal{A} = ((a_{i,j}))$ called a “ratio scale matrix” (RSM). A direct computation proves that if $\mathbf{a}^N \in \mathcal{CN}^N$ and \mathbf{w}^N is \mathcal{A} ’s eigenvector associated with its sole positive eigenvalue (Perron–Frobenius theorem), then \mathbf{w}^N satisfies all of the above conditions. The problem is to handle $\mathbf{a}^N \notin \mathcal{CN}^N$. A natural approach is to replace \mathbf{a}^N with its “closest” \mathcal{CN}^N point. Several such methods are described by Colany and Kress (1993) demonstrating their different outcomes. These differences are explained below.

The Crawford–Williams approach differs; for any \mathbf{a}^N , it uses the geometric mean to define

$$w_j = (a_{j,1}a_{j,2} \dots a_{j,N})^{\frac{1}{N}}, \quad j = 1, \dots, N. \quad (7.52)$$

To motivate this choice with the normalization $w_1w_2 \dots w_N = 1$, if Equation (7.50) holds, then

$$(a_{j,1}a_{j,2} \dots a_{j,N})^{\frac{1}{N}} = \left(\frac{w_j}{w_1} \times \frac{w_j}{w_2} \times \dots \times \frac{w_j}{w_N} \right)^{\frac{1}{N}} = \left(\frac{w_j^N}{w_1w_2 \dots w_N} \right)^{\frac{1}{N}} = w_j.$$

For $\mathbf{a}^N \in \mathcal{CN}^N$, the eigenvector and geometric mean approaches define (up to a scale multiple) the same \mathbf{w}^N . However, if $\mathbf{a}^N \notin \mathcal{CN}^N$, they can disagree. The question is to determine which approach provides consistent answers.

Fortunately, by use of Sections 7.4.1–7.4.4, this issue can be immediately resolved. This is because the additive and multiplicative structures are connected by the group isomorphism

$$F : \mathcal{BO}^N \rightarrow \mathcal{SC}^N \text{ defined by } a_{i,j} = e^{d_{i,j}}, 1 \leq i \leq j \leq N. \quad (7.53)$$

Thus $F(\mathbf{d}^N) = \mathbf{a}^N = (a_{1,2}, a_{1,3}, \dots, a_{1,N}; a_{2,3}, \dots; a_{N-1,N})$. Equation (7.49) is satisfied because $1 = F(0) = F(d_{i,j} + d_{j,i}) = F(d_{i,j})F(d_{j,i}) = a_{i,j}a_{j,i}$, so $a_{i,j} = 1/a_{j,i}$ and $a_{j,i} = 1$.

Similarly, the Equation (7.31) strong transitivity condition is identified with the Equation (7.51) consistency expression, which means (Equation (7.53)) that \mathcal{ST}^N and \mathcal{CN}^N are isomorphic. By being isomorphic, *conclusions developed in Sections 7.4.1–7.4.4 transfer to explain what happens with ratio scaling models*. As an example, BAR and $\mathbf{w}^N \in \mathcal{CN}^N$ values are identified:

$$\bar{b}_i = \sum_{j=1}^N d_{i,j} = \sum_{j=1}^N \ln(a_{i,j}) = \ln(\prod_{j=1}^N a_{i,j}). \quad (7.54)$$

Theorem 7.23 *If $\mathbf{a}^N \in \mathcal{CN}^N$ (so Equation (7.51) is satisfied), then $\bar{b}_i = N \ln(w_i) - \ln(C)$, and, conversely, $w_i = [Ce^{\bar{b}_i}]^{\frac{1}{N}}$, $i = 1, \dots, N$, where $C = w_1 w_2 \dots w_N$.*

The proof follows immediately from Equations (7.51)–(7.54).

7.5.4.1 Problems with Projections

Implicit in using projections is that the data space has a decomposition where the cleansed version is directly (i.e., orthogonally located) below the given noisy data. This is true with \mathcal{BO}^N as proved with its theoretical decomposition Theorem 7.9. With the \mathcal{SC}^N structure, however, normal projections are inappropriate.

One reason is that \mathcal{CN}^N is a non-linear $(N - 1)$ -dimensional manifold. Indeed, holding $a_{i,k}$ fixed in Equation (7.51) requires the admissible $a_{i,j}$ and $a_{j,k}$ to define a hyperbola similar to that depicted in Figure 7.8. As indicated in Figure 7.8, this non-linearity means that, in general, the “nearest” \mathcal{CN}^N point is *not* the cleansed version of a given \mathbf{a}^N .

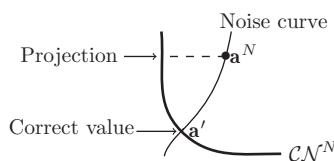


Figure 7.8 Problems with projections.

Projections require an understanding of the “noise” structure. To describe this with $N = 3$, Equation (7.36) (by using $\ln(f)$ coefficients rather than f) can be expressed as

$$\mathbf{d}^3 = \sum_{j=1}^2 \ln(\beta_j) \mathbf{B}_j^3 + \ln(c) \mathbf{C}_{\{1,2,3\}}^3, \quad (7.55)$$

where $\beta_1, \beta_2, c > 0$. Thus, for instance, $d_{1,2} = \ln(\beta_1) - \ln(\beta_2) + \ln(c) = \ln(\frac{c\beta_1}{\beta_2})$. In this manner, $\mathbf{d}^3 = (\ln(\frac{c\beta_1}{\beta_2}), \ln(\frac{\beta_1}{c}); \ln(c\beta_2))$, so $\mathbf{a}^3 \in \mathcal{SC}^3$ becomes

$$\mathbf{a}^3 = (\frac{c\beta_1}{\beta_2}, \frac{\beta_1}{c}; c\beta_2); \quad \beta_1, \beta_2, c \in (0, \infty). \quad (7.56)$$

As each \mathbf{d}^3 has a unique Equation (7.55) representation, each \mathbf{a}^3 has a unique Equation (7.56) representation.

The non-linear Equation (7.56) represents the “cyclic noise line” (depicted in Figure 7.8); it describes a natural progression from consistency ($c = 1$) through all $c \neq 1$ inconsistency levels.

Theorem 7.24 *For $\mathbf{a}^3 \in \mathcal{SC}^3$, unique $\mathbf{a}' \in \mathcal{CN}^3$, β_1, β_2 , and c define*

$$\mathbf{a}^3 = (a'_{1,2}c, \frac{a'_{1,3}}{c}; a'_{2,3}c), \quad a'_{1,2} = \frac{\beta_1}{\beta_2}, a'_{1,3} = \beta_1, a'_{2,3} = \beta_2. \quad (7.57)$$

Namely, $\mathbf{a}^3 \notin \mathcal{CN}^3$ is a distorted version of a unique $\mathbf{a}' \in \mathcal{CN}^3$ caused by the cyclic noise component c . The noise (c terms) extends an $\mathbf{a}' \in \mathcal{CN}^3$ to generate a non-linear curve in \mathcal{SC}^3 . (These Equation (7.57) values come from the coordinate representation of $\mathbf{d}^3 = F^{-1}(\mathbf{a}^3)$.)

To find the appropriate \mathbf{w}^3 for a given \mathbf{a}^3 , the analysis requires finding the appropriate $\mathbf{a}' \in \mathcal{CN}^3$, which (Equation (7.57), Figure 7.8) involves non-linearities. No projection method I have investigated comes even close to capturing this structure, which explains the inconsistencies in the various choices. In other words, projection approaches cannot reflect the theoretical structure of removing noise. As indicated below, everything extends to $N \geq 3$.

Even more; explanations for cyclic outcomes follow from Equation (7.57).

Corollary 7.1 *For $\mathbf{a}' \in \mathcal{CN}^3$, let \mathbf{a}_c^3 be the associated Equation (7.57) term for c . There exists a $c_m < 1$ and a $c_M > 1$ so that if $0 < c < c_m$, the \mathbf{a}_c^3 ranking (defined by $a_{i,j}$ values) is a cycle, while if $c > c_M$, the \mathbf{a}_c^3 ranking is the opposite cycle.*

Let the \mathbf{a}' ranking be $A_i > A_j > A_k$ with distinct $a_{u,v}$ values. Because $a_{i,j}, a_{j,k}, a_{i,k} > 1$, consistency requires that either $a_{i,j}$ or $a_{j,k}$ has the smallest value; as the c value varies, four sets of strict rankings emerge: the original transitive ranking, a second transitive ranking where the pair with the smallest $a_{u,v} > 1$ value is reversed, and the two sets of cyclic rankings.

Proof The proof of this statement (along with the c_m , c_M , and transition values) follows directly from elementary algebra and Equation (7.57). \square

Everything extends to $N \geq 3$. Some differences are caused by the higher dimension of \mathcal{CY}^N , e.g., a non-linear coordinate system for $N = 4$ with $t\mathbf{C}_{(1,2,3,4)}^4 + u\mathbf{C}_{(1,3,2,4)}^4 + v\mathbf{C}_{(1,3,4,2)}^4$ is

$$\mathbf{a}^4 = (a'_{1,2} \frac{t}{v}, \quad a'_{1,3}uv, \quad a'_{1,4} \frac{1}{tu}; \quad a'_{2,3} \frac{t}{u}, \quad a'_{2,4} \frac{u}{v}; \quad a'_{3,4}tv) \quad (7.58)$$

and the matrix

$$B = \begin{pmatrix} 1 & \frac{w_1 t}{w_2 v} & \frac{w_1}{w_3} uv & \frac{w_1}{w_4 tu} \\ \frac{w_2 v}{w_1 t} & 1 & \frac{w_2 t}{w_3 u} & \frac{w_2 u}{w_4 v} \\ \frac{w_3}{w_1} \frac{1}{uv} & \frac{w_3 u}{w_2 t} & 1 & \frac{w_3}{w_4} tv \\ \frac{w_4}{w_1} tu & \frac{w_4 v}{w_2 u} & \frac{w_4}{w_3} \frac{1}{tv} & 1 \end{pmatrix} \quad (7.59)$$

which defines a highly non-linear three-dimension “noise” surface.

7.5.4.2 Returning to the Geometric Mean

The principles needed to use a projection do not apply in ratio-scaling settings. First, the underlying space of desired outcomes is highly non-linear, so “closest” points, or projections, need not be the desired outcome. Second, as captured by Equations (7.57) and (7.58) (and Figure 7.8) the noise structure has a highly non-linear structure. To find the associated \mathcal{CN}^N point for \mathbf{a}^N , this line must be traced to find its intersection with \mathcal{CN}^N ; doing so is not compatible with normal projection methods.

Finding the associated \mathcal{CN}^N point for \mathbf{a}^N appears to be highly complicated. Fortunately, it is not; answers come by transferring information about how the additive structure of BAR values drop cyclic terms. With Equation (7.53), cyclic error terms should drop with multiplication. Indeed, defining $\bar{a}_j = \prod_{k=1}^N a_{j,k}$, it follows from Equations (7.56) and (7.58) (in the Equation (7.58) matrix by multiplying terms in the j th row) that all cyclic terms cancel, leaving a multiple of $(w_j)^N$. Thus, an easy way to eliminate noise effects and obtain a \mathbf{w}^N that is supported by theory is to use the geometric mean approach. Even more; by using F to translate Theorem 7.10 into this multiplicative setting, it follows that the geometric mean is essentially the only way to do so. (Some changes are needed; e.g., in Theorem 7.10, the condition that \mathcal{CY}^N is in the kernel becomes that the method does not depend on cyclic terms.)

Much more can be stated. For instance, it follows from Theorem 7.12 that *all* of the above methods will suffer changes in rankings by removing alternatives. As with Theorem 7.12, the problem is caused by the cyclic noise. Resolutions are the same; instead of perpetuating complexities caused by cyclic noise components, replace the data with its cleansed \mathcal{CN}^N form. Everything follows directly from the linear structure.

7.6 Summary

For centuries, aggregation methods have suffered difficulties. While clever approaches have been created to avoid some complexities, other problems could arise. Adopting a theoretical approach finally leads to solutions.

While open problems in a couple of disciplines have been resolved by using the approaches described here, the methodology is not unusual; it is to find an appropriate coordinate system for the data – and justify why it is the “correct one.” Without question, this theoretical approach extends to other disciplines and systems.

References

- Arrow, K. J. (1951). *Social choice and individual values*. New York, NY: Wiley.
- Bargagliotti, A., & Saari, D. G. (2010). Symmetry of nonparametric statistical tests on three samples. *Journal of Mathematics and Statistics*, 6(4), 395–408.
- Black, D. (1958). *The theory of committees and elections*. London: Cambridge University Press.
- Bhapkar, V. P. (1961). A nonparametric test for the problem of several samples. *Annals of Mathematical Statistics*, 32, 1108–1117.
- Bhapkar, V. P., & Deshpande, J. V. (1968). Some nonparametric tests for multisample problems. *Technometrics*, 10, 578–585.
- Colany, B., & Kress, M. (1993). A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices. *European Journal of Operational Research*, 69, 210–220.
- Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris.
- Crawford, G., & Williams, C. (1985). A note on the analysis of subjective judgement matrices. *Journal of Mathematical Psychology*, 29, 387–405.
- Crawford, G. B. (1987). The geometric mean procedure for estimating the scale of a judgement matrix. *Mathematical Modelling*, 9, 327–334.
- Gibbard, A. (1969). *Social choice and the Arrow conditions*. Unpublished manuscript; University of Michigan.
- Haunsperger, D. (1992). Dictionaries of paradoxes for statistical tests on k samples. *Journal of the American Statistical Association*, 87, 149–155.
- Jessie, D., & Saari, D. G. (2016). From the Luce choice axiom to the quantal response equilibrium, *Journal of Mathematical Psychology*, 75, 3–9.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47, 583–612.
- Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.
- Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15, 215–233.
- Maskin, E., & Sen, A. (2016). How majority rule might have stopped Donald Trump. Opinion article. *New York Times*, April 28, 2106.
- McKelvey, R., & Palfrey, T. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10, 6–38.
- Nakamura, K. (1975). The core of a simple game with ordinal preferences. *International Journal of Game Theory*, 4, 95104.
- Nakamura, K. (1978). The voters in a simple game with ordinal preferences. *International Journal of Game Theory*, 8, 5561.
- Regenwetter, M., Kim, A., Kantor, A., & Ho, M. H. (2007). The unexpected empirical consensus among consensus methods. *Psychological Science*, 18, 559–656.

- Saari, D. G. (1989). A dictionary for voting paradoxes. *Journal of Economic Theory*, 48, 443–475.
- Saari, D. G. (1992a). The aggregate excess demand function and other aggregation procedures. *Economic Theory*, 2, 359–388.
- Saari, D. G. (1992b). Millions of election rankings from a single profile. *Social Choice and Welfare*, 9, 277–306.
- Saari, D. G. (1995a). *Basic geometry of voting*. New York, NY: Springer.
- Saari, D. G. (1995b). A chaotic exploration of aggregation paradoxes. *SIAM Review*, 37, 37–52.
- Saari, D. G. (1999). Explaining all three-alternative voting outcomes. *Journal of Economic Theory*, 87, 313–355.
- Saari, D. G. (2000a). Mathematical structure of voting paradoxes 1: Pairwise vote. *Economic Theory*, 15, 1–53.
- Saari, D. G. (2000b). Mathematical structure of voting paradoxes 2: Positional voting. *Economic Theory*, 15, 55–101.
- Saari, D. G. (2001a). Analyzing a “nail-biting” election. *Social Choice and Welfare*, 18, 415–430.
- Saari, D. G. (2001b). *Decisions and elections: Explaining the unexpected*. New York, NY: Cambridge University Press.
- Saari, D. G. (2005). The profile structure for Luce’s choice axiom. *Journal of Mathematical Psychology*, 49, 226–253.
- Saari, D. G. (2008). *Disposing dictators, demystifying voting paradoxes*. New York, NY: Cambridge University Press.
- Saari, D. G. (2014a). A new way to analyze paired comparison rules. *Mathematics of Operations Research*, 39, 647–655.
- Saari, D. G. (2014b). Unifying voting theory from Nakamura’s to Greenberg’s Theorems. *Mathematical Social Sciences*, 69, 1–11.
- Saari, D. G. (2016). Basis for binary comparisons and non-standard probabilities. Special issue of *Philosophical Transactions of the Royal Society A*, Quantum probability and the mathematical modeling of decision-making, 374(2058), 20150103
- Saari, D. G., & Tataru, M. (1999). The likelihood of dubious election outcomes. *Economic Theory*, 13, 345–363.
- Saaty, T. (1980). *The analytic hierarchy process: Planning, priority setting, resource allocation*. New York, NY: McGraw-Hill.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Ward, B. (1965). Majority voting and the alternative forms of public enterprise. In J. Margolis (ed.), *The public economy of urban communities* (pp. 112–126). Baltimore, MD: Johns Hopkins University Press.
- Wang, Z., Solloway, T., Shiffrin, R., & Busemeyer, J. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Science*, 111(26), 9431–9436.
- Yellott, J. I., Jr. (1997). The relationship between Luce’s Choice Axiom, Thurston’s theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, 15, 109–144.

8 Categorization Based on Similarity and Features: The Reproducing Kernel Banach Space (RKBS) Approach

Jun Zhang and Haizhang Zhang

8.1	Introduction	323
8.1.1	Usefulness of Categories	324
8.1.2	Extant Psychological Models and Issues	325
8.1.3	An Emerging Unifying Framework	327
8.1.4	Plan of Chapter	328
8.2	Mathematical Preliminaries	329
8.2.1	Vector Spaces and Functionals	329
8.2.1.1	Linear Functionals and Norms	330
8.2.1.2	Inner Products and Hilbert Spaces	331
8.2.1.3	Semi-inner Products and Banach Spaces	333
8.2.1.4	Duality Mapping and Generalized Semi-inner Products	334
8.2.1.5	Bases in Hilbert and Banach Spaces	335
8.2.1.6	Frames in Hilbert and Banach Spaces	336
8.2.2	Function Spaces and Reproducing Kernels	337
8.2.2.1	Reproducing Kernel Hilbert Spaces (RKHS)	338
8.2.2.2	Reproducing Kernel Banach Spaces (RKBS)	339
8.2.2.3	Operator-valued Reproducing Kernels	339
8.2.2.4	Measurable Functions as Vector Spaces: $L^p(\mathcal{X}), M(\mathcal{X}), C_0(\mathcal{X})$, etc.	341
8.2.2.5	Integral Operators and Spectral Methods	342
8.3	Learning as Regularization: Problem Formulation	343
8.3.1	Statistical Learning Theory	343
8.3.1.1	Assumptions on Input and Output Spaces	344
8.3.1.2	Learning as an Optimization Problem	344
8.3.1.3	Risk of Generalization and Error Decomposition	344
8.3.2	Regularization by Different Norms	345
8.3.2.1	Norms to Enforce Smoothness	346
8.3.2.2	Norms to Enforce Sparsity	346

8.3.2.3	Different Norms on Kernelized Functions	347
8.4	Learning with Reproducing Kernels: Solution Concepts	347
8.4.1	Power of Reproducing Kernels	347
8.4.1.1	Representer Theorem	347
8.4.1.2	Feature Spaces and Feature Maps	348
8.4.1.3	Kernel Trick	349
8.4.2	Sampling via Reproducing Kernels	350
8.4.2.1	Shannon Sampling	350
8.4.2.2	Sample-based Hypothesis Space with ℓ^1 -Norm	350
8.4.3	Kernel Method as a Unifying Feature and Similarity	351
8.4.3.1	Dictionary Learning and Feature Selection	352
8.4.3.2	Maximal Margin Classifier and SVM	353
8.4.3.3	Kernel Learning and Vector-valued Maps	353
8.5	Psychological Underpinnings of Kernel Methods	354
8.5.1	Computational Models of Human Categorization	354
8.5.1.1	Exemplar Models	355
8.5.1.2	Prototype Models	356
8.5.1.3	Decision-bound Models	357
8.5.1.4	Connectionist Models	358
8.5.1.5	Bayesian Models	359
8.5.2	Unified Account of Psychological Categorization Models	361
8.5.2.1	Relations Among Various Categorization Models	361
8.5.2.2	Unified Account by Reproducing Kernels	362
8.5.2.3	Shepard Kernel and Multidimensional Input Scaling	363
8.5.2.4	Incorporating Attention into Kernel Methods	364
8.5.3	Challenges for Kernel Methods	365
8.5.3.1	Similarity vs. Dissimilarity	365
8.5.3.2	Flexible Feature Representation	366
8.6	Summary: Unifying Feature and Similarity by RKBS	367
	Acknowledgement	368
	References	368

8.1 Introduction

Categorization refers to the process by which discriminably different things are judged as belonging to groups that are treated as being equivalent in some regard. The ability to categorize or classify objects, images, events, actions, etc., is a hallmark of adaptive intelligence. It allows animals and humans to robustly extract regularity and patterns in environmental inputs without much prior knowledge or assumptions about their structures and interrelations, thereby reducing the complexity of mental computation to a manageable degree. Categorization arises as a result of a flexible search for structure of things in the environment which consists of patterns of correlated features forming natural chunks (clusters). Category learning is the process of inferring the structure of categories from a set of

stimuli labeled as belonging to those categories; the knowledge acquired through this process can ultimately be used to make decisions about how to categorize new stimuli.

To an adaptive agent, categorization (as it is called in cognitive psychology) or classification and clustering (as it is called in machine learning) provides an internal model of the environment, with dimensionality of the “state space” greatly reduced compared with how inputs are specified; this solves the bottleneck problem facing reinforcement learning, which deals with the agent’s optimal action planning and sequential decision-making. Solution of the categorization problem, in combination with the now well-understood solution to reinforcement learning, will pave the way for designing artificial intelligence with autonomous and adaptive planning capabilities.

In this chapter, we review a theoretical framework of categorization that is both strongly motivated by cognitive psychology and rigorously grounded in mathematics. This is the framework of regularized learning under supervision using finite samples. We provide a unified treatment of *2-norm based reproducing kernel method for regularized learning* with *1-norm based sparsity method for feature selection*, built upon Banach space formulation of the learning problem. The mathematical tool for our exposition is the reproducing kernel Banach space (RKBS) – an RKBS is equipped with a semi-imperfect discrimination operation, which specializes to the imperfect discrimination operation for the case of Hilbert spaces. The RKBS framework provides a potentially unifying framework for the analysis of similarity and feature representation in categorization.

8.1.1 Usefulness of Categories

Categories serve two primary functions (see Smith, 1995): (1) they enable efficient storage and retrieval of information, and relieve us from the burden of keeping track of every individual item we encounter; (2) they promote inferences and extend our knowledge beyond past experiences into the future, allowing us to make predictions that guide behavior. Categories serve not only to organize the knowledge we have already acquired, but also to guide our expectations through inductive reasoning. Induction is the mental capacity to extend knowledge to novel instances, for example, inferring that a newly encountered mushroom is poisonous on the basis of past encounters with other poisonous mushrooms. Inductive inference is one of the most important utilities that come with categorization.

The ability to categorize arises very early ontogenetically. Young children are able to use categories to support inductive inferences even when category membership of objects conflicts with their appearances, whereas category labels do not generally affect children’s perceptual similarity judgments (Noles & Gelman, 2012). Young children’s category-based inferences attained the status of what Gelman dubbed as “psychological essentialism” – children readily infer properties that concern internal features and non-visible functions from one category member to

another, and their inferences rely more on category membership information than perceptual similarity information (Gelman, 2004).

There has accumulated a large body of empirical literature on human categorization. Rosch's (e.g., Rosch, 1973; Rosch & Mervis, 1975) seminal studies of natural object categories led to the commonly held view that people learn about and use hierarchically organized categories, with entry-level categories possessing meaningful sub-structures with correlated features and super-structures that they contribute towards. Feature and category label are two sides of the coin of a hierarchical scheme – the fundamentally probabilistic nature of the environment leads to the creation of category labels for objects with overlapping features based on some similarity function defined on some feature space. However, perceptual similarity alone does not account for differences in categories, although our perceptual system clearly has evolved to facilitate categorization. As concepts and categories serve (sometimes multiple) functions and goals, the categorization system must be flexible enough to both assimilate the necessary underlying structure and discover or even create that structure.

The importance of solving the categorization puzzle can be appreciated in connection with reinforcement learning. Whereas reinforcement learning (both model-based and model-free variants) enables optimal planning and sequential decision-making for an agent, category induction from finite samples will construct a compact representation of environmental state space. Categorization is both a hallmark of adaptive intelligence and the bottleneck problem affecting wider applications of the otherwise successful reinforcement learning algorithms.

8.1.2 Extant Psychological Models and Issues

There is a long tradition of empirical study and theoretical construction of mathematical models of human categorization by cognitive psychologists. Existing psychological models of categorization deal both with the process and the mechanism of various aspects of categorization. On the *normative* side, there is the rational theory of categorization (Anderson, 1990, 1991) which essentially treats category learning as non-parametric Bayesian inference about unknown probability distributions. This approach is followed by the SUSTAIN model (Love, Medin, & Gureckis, 2004; Vanpaemel & Storms, 2008), treating categories as mixture distributions over clusters, and becomes fully developed by Griffiths and associates (Austerweil & Griffiths, 2011, 2013; Sanborn, Griffiths, & Navarro, 2010) using elaborative statistical tools about sampling. On the *representational* side, there are basically two camps advocating opposing views about memory representations that support categorization. The prototype model (Reed, 1972; Smith & Minda, 1998) starts with the assumption that humans supposedly “average” their experience of encountering various exemplars of a category to create a prototype that is most typical for each category. This common-sense model was challenged by the class of exemplar model (Medin & Schaffer, 1978; Nosofsky, 1986), which claims that

humans store many exemplars of the same category in memory and those exemplars are simultaneously retrieved and compared against upon encountering a new input to be categorized. The debate between prototype model and exemplar model is how categories are represented in the memory system, whether as summary statistics (with mean and covariance structure) or as individual instances. On the *mechanistic* side, there are also different conceptualizations of how human category decisions are arrived at. The decision-bound model (Ashby & Gott, 1988; Ashby & Perrin, 1988) argues that humans represent stimuli in a multidimensional vector space that can be partitioned into various regions representing equivalent classes (with distinct category labels). Furthermore, the boundary of these partitioned regions can be explicitly learnt to form “decision boundaries.” The connectionist/neural network model (Gluck & Bower, 1988; Kruschke, 1992), on the other hand, argues that input stimuli are represented in a parallel-distributed fashion that cannot be simply conceptualized as a vector space, and categorization is the result of non-linear network computation where non-linearity is crucial for both its representational power and classification accuracy. So, while the debate between prototype/exemplar models reflects contrasting theoretical assumptions about the representation and memory requirements for category structures, the debate between decision-bound/distributive network models reflects contrasting theoretical commitments about the architecture and implementation styles for the process and dynamics of category decisions.

The set of common issues to be addressed by computational models of human categorization include:

- (i) The role of exemplar versus prototype in categorization: How is category structure maintained in memory? Are all exemplars equally important, or does the typical exemplar (prototype) play a special role such as summary statistics?
- (ii) The relationship between feature and category label: How is feature representation created? Does category label serve as just another feature?
- (iii) The architecture for category decisions: Are there clear boundaries in some multidimensional feature space that delineate one category from another? Do category decisions arise from connectionist/neural network style non-linear computation reflecting emerging properties that may not be mathematically trackable? How do rule-based versus boundary-based computation schemes reconcile with distributive representation?
- (iv) The importance of similarity and generalization during categorization: Is Sheperd’s universal law of generalization an adequate expression of similarity? Is similarity computation feature-based or exemplar-based?

A synthesis of the above-mentioned psychological models, all receiving various degrees of empirical support, would not only provide a satisfactory account of human categorization performance but greatly aid the search for a universal algorithm of the learning-from-sample problem in machine learning. To do so, one needs to combine the strengths of exemplar-based versus feature-based

approaches, and integrate bound-based versus cluster-based approaches to explain how mechanisms (e.g., attention, similarity) at one level translate to another.

8.1.3 An Emerging Unifying Framework

Over the past decades, the field of statistical machine learning has produced a suite of power tools to tackle the problem of supervised and semi-supervised classification based on finite samples. The mathematical framework is the so-called “regularized loss minimization,” which formulates the goal of classification as building an optimal model (i.e., input–output mapping) capable of generalizing beyond given sample data. The general philosophy is to treat categorization as an “ill-defined” or underconstrained inverse problem – there can be multiple schemes of category structures (“output”) that can account for sample data (“input”), so there can be no unique solution of an induced classification scheme that is compatible with finite sample data. Instead, one looks for an “optimal” solution, with optimality phrased in terms of the goals of categorization. A key insight in the computational approach to categorization is that the input–output mapping should be evaluated based on its performance not only on known samples but also on unseen data. Technically, when misclassification is quantified as “loss,” the objective for the learning algorithm is to minimize such loss by choosing a classifier with a modest amount of complexity (“regularized”) while generalizing well from seen to unseen samples.

There are two main approaches in regularized learning, however, with entirely different emphasis and mathematical underpinning: one based on *reproducing kernel methods* and the other based on *sparsity methods*. Although both adopt the same principle of regularization, i.e., balancing the competing needs for goodness-of-fit and for simplicity of the optimal classifier, they differ with respect to generalization mechanism and feature representation. On the one hand, the paradigm of kernel methods assumes that classifiers live in a function space that is a Hilbert space with a reproducing kernel. The reproducing kernel, or kernel in short, guarantees the existence of (what can be an extremely non-linear) mapping from essentially arbitrary inputs to their feature representation, in service of, say, hyperplane-based classification (e.g., maximal margin methods). Although representation of features as a high-dimensional vector space is conceptually invoked, such a paradigm does not rely on explicit construction of the feature space. Instead, a kernel function is invoked to measure similarity of objects in the input space, such that their feature representation is often “bypassed” – this is known in the research community as the “kernel trick.” On the other hand, the developments in Lasso in statistics (Tibshirani, 1996) and compressive sensing in signal processing (Candès, Romberg, & Tao, 2006), emphasize the importance of sparse representation for coding and classification using properly selected bases. As is now well-understood, it is the ℓ^1 -norm (rather than the L^2 -norm of a Hilbert space) that enforces the sparsity solution (as surrogate to the computationally intractable ℓ^0 regularization problem) and enables optimal feature selection. However, when

regularization learning is formulated on the Banach space of functions with L^1 -norm (as opposed to the sequence space ℓ^1 where bases are already chosen), no representer-type theorem is known.

The connection and interaction between similarity-based generalization (through L^2 -regularization) and the sparsity-based feature selection (through ℓ^1 -regularization), both of which are reasonably well-understood, lie at the crux of solving the categorization problem. From cognitive psychology, the literature is quite clear that both similarity-based generalization (captured by the L^2 -based kernel method) and feature-based attention (modeled as ℓ^1 -based sparse feature selection) are at play during category learning. Although the relevance of the RKHS framework for exemplar and prototype or perceptron models has been discussed abstractly (Jäkel, Schölkopf, & Wichmann, 2008a,b, 2009), and its potential for neural representations of feature/object dichotomy has been suggested (Riesenhuber & Poggio, 1999; Smale *et al.*, 2010), the details have yet to be laid out.

Besides features and similarity, another important cognitive construct in categorization and category learning/induction is that of attention. Existing psychological theories of attention in learning, as reviewed below, can be interpreted as modifying the kernel by rescaling stimulus representations either in the input space (Nosofsky, 1986; Sutherland & Mackintosh, 1971) or in the feature space (Kruschke, 2001; Mackintosh, 1975). Kernel learning (Lanckriet *et al.*, 2004) investigates the relationship between various forms of representation learning, in the context of convex optimization of mixture kernels (Argyriou, Micchelli, & Pontil, 2005; Micchelli & Pontil, 2005a) or multiple tasks (Evgeniou, Micchelli, & Pontil, 2005). Attention acting on the feature space may induce a convex family of kernels, but attention acting on the input space generally does not. Therefore, the goal of optimizing the kernel through attention to input space (as in exemplar models of categorization) is of particular interest. There is a need for separate mathematical analyses of attention on exemplars versus attention on feature dimensions.

8.1.4 Plan of Chapter

The plan of the rest of this chapter is as follows. In Section 8.2, we review fundamental mathematical concepts and tools of reproducing kernel methods. Starting from vector spaces, we introduce the notions of linear functional, norm, duality mapping, inner product and its generalization semi-inner product (Section 8.2.1). When considering function spaces as special examples of vector spaces, the idea of the reproducing kernel is then introduced to link a function with its evaluation (Section 8.2.2). In Section 8.3, the mathematical problem of learning with finite samples is formulated. This is done by first revisiting statistical learning theory and the need for regularization (Section 8.3.1), and the various norms used as regularizers (Section 8.3.2). Section 8.4 then proceeds by reviewing solution concepts afforded with the reproducing kernel methods. They include the celebrated

“representer theorem” and the widely used “kernel trick” (Section 8.4.1), sampling methods (Section 8.4.2), and maximal margin classifier, feature selection, and vector-valued functions as feature maps (Section 8.4.3). Section 8.5 deals with psychological underpinnings of kernel methods. We first provide a comprehensive review of various psychological models of human categorization (Section 8.5.1), and then attempt to cast these models in a unified framework using the language of kernel methods (Section 8.5.2), with challenges also discussed (Section 8.5.3). Finally, our chapter closes with a short summary (Section 8.6).

8.2 Mathematical Preliminaries

In this section, we review the mathematical background underlying the reproducing kernel method that is popular in contemporary machine learning. We first review a basic but important mathematical object, namely, “vector space.” Then, we study functions defined on vector spaces, known as “functionals,” and the duality mapping between the space of continuous linear functionals and the original vector space; such duality mapping is linked to inner product or semi-inner product defined on a Hilbert or a Banach (vector) space, respectively. Next, we focus on function spaces as a particular kind of vector space, and the evaluation functional as a particular kind of linear functional. This leads to the notion of “reproducing kernels,” which plays a central role in formulating the regularization problem in machine learning.

8.2.1 Vector Spaces and Functionals

A vector space V over the field \mathbb{C} of complex numbers is a set equipped with the following two operations that are closed for all $u, v \in V, c \in \mathbb{C}$:

1. *vector addition*: $(u, v) \rightarrow u + v \in V$;
2. *scalar multiplication*: $(c, u) \rightarrow cu \in V$.

The vector addition operation is required to be commutative $u + v = v + u$, associative $w + (u + v) = (w + u) + v$, and to possess an identity $0 \in V$, namely $0 + u = u + 0 = u$ for all $u \in V$, such that every vector $u \in V$ has an additive inverse $-u \in V$, namely $u + (-u) = 0$. Essentially, these requirements about the vector addition operation + make $(V, +)$ an Abelian group. The scalar multiplication operation, on the other hand, “couples” the Abelian group structure with the scalar field to provide richer structures of V . First, the scalar multiplication is associative and distributive with respect to vector addition. That is, there hold for all $u, v \in V$ and $a, b \in \mathbb{C}$ that

$$a(bu) = (ab)u, (a + b)u = au + bu, a(u + v) = au + av.$$

Second, the unity $1 \in \mathbb{C}$ in the field satisfies $1u = u$ for all $u \in V$. Note that a number field (in this case \mathbb{C}) must be associated with a vector space V – one

cannot speak of a vector space without specifying the associated field. However, the field can be any general number field, including the field \mathbb{R} of real numbers.

A familiar example of vector space is \mathbb{C}^n (or \mathbb{R}^n), where every vector is represented by an array of n complex (or real) numbers, addition is componentwise, and scalar multiplication is field multiplication on each of these numbers separately.

8.2.1.1 Linear Functionals and Norms

Over the vector space V , one can consider mappings from V to \mathbb{C} , that is, taking any vector in V as input to yield a number $c \in \mathbb{C}$ as output. Each such mapping, denoted as T , is called a *functional*, the totality of which we denote as \mathbb{C}^V .

An important property of the space \mathbb{C}^V of functionals (with vectors in V as prescribed inputs) is that \mathbb{C}^V itself is a vector space. In particular, linear combinations $c_1T_1 + c_2T_2$ of two functionals T_1, T_2 are functionals (where $T_1, T_2 \in \mathbb{C}^V$ and c_1, c_2 are scalars). Among all functionals, there is a particular subset called *linear functionals*, i.e., the map T is linear with respect to the *input* vectors u, v in V : $T(u+v) = T(u) + T(v)$. The set of all linear functionals forms a vector space, a subspace of \mathbb{C}^V .

Given a vector space V over \mathbb{C} , an important functional is the *norm* on V , usually denoted as $\|\cdot\|_V$ or simply $\|\cdot\|$. A norm is a non-linear functional which maps an element of V to a non-negative real value \mathbb{R}_+ such that

- (i) *positivity*: $\|u\|_V \geq 0$ for all $u \in V$ and $\|u\|_V = 0$ if and only if $u = 0$,
- (ii) *homogeneity*: $\|cu\|_V = |c|\|u\|_V$ for all $u \in V$ and $c \in \mathbb{C}$,
- (iii) *triangle inequality*: $\|u+v\|_V \leq \|u\|_V + \|v\|_V$ for all $u, v \in V$.

A vector space endowed with a norm is called a *normed vector space*.

Any norm $\|\cdot\|_V$ on a vector space V induces a metric $d(\cdot, \cdot)$ on V by

$$d(u, v) = \|u - v\|_V, \quad u, v \in V.$$

It can be verified, by the three defining properties of a norm, that $d(u, v)$ is non-negative, symmetric, and satisfies the triangle inequality

$$d(u, v) \leq d(u, w) + d(w, v).$$

Thus, $d(u, v)$ qualifies as a metric and thus can induce a topology on V . Under this norm-induced topology, the convergence of a sequence to some point is equivalent to their metric distance vanishing in the limit.

For a normed vector space V , we can speak of continuity of any linear functional T mapping V to \mathbb{C} : continuity is with respect to the norm-induced topology on V . The set of all continuous linear functionals on V is denoted as V^* ; it is a vector subspace of \mathbb{C}^V , and is called the “dual space” or simply the “dual” of V . It can be equipped with the so-called “dual norm” (induced from the norm $\|\cdot\|_V$ on V):

$$\|T\|_{V^*} := \sup_{u \in V, u \neq 0} \frac{|T(u)|}{\|u\|_V}.$$

(One can prove that $\|T\|_{V^*}$ as defined indeed satisfies the three axioms of a norm.)

When $\|T\|_{V^*} \leq C$ where C is some positive real number, then we say that T is a bounded functional. When $\|T\|_{V^*} < +\infty$, then T is continuous. Equivalently, $\lim_{n \rightarrow \infty} T(u_n) = 0$ whenever u_n converges to $0 \in V$. The Hahn–Banach theorem, a fundamental result in functional analysis, states that every continuous linear functional on a subspace of the input vector space V can be extended in a norm-preserving way to the whole space V . Hence, for each $u \in V$ there exists a continuous linear functional $T \in \mathbb{C}^V$ such that

- (i) $|T(u)| = \|T\|_{V^*} \|u\|_V$;
- (ii) $\|T\|_{V^*} = \|u\|_V$.

We call any linear functional satisfying (i) a “dual element” or simply “dual” of u (or *with respect to u*) and usually denote it as u^* . When u maps to u^* , cu maps to $\bar{c}u^*$ for any $c \in \mathbb{C}$. This is to say, vectors in the same direction in V map to identical direction in the dual space V^* . When (ii) is further satisfied, u^* is called the “canonical dual” of u . In general, one can require instead of (ii): $\|T\|_{V^*} = \gamma(\|u\|_V)$ for a strictly monotone function $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ that satisfies $\gamma(0) = 0$; $\lim_{t \rightarrow \infty} \gamma(t) = \infty$; the function γ is called a “gauge function.”

The canonical dual $u^* \in V^*$ of $u \in V$ is unique when the norm $\|\cdot\|_V$ is uniformly convex. When the canonical dual of each $u \in V$ is chosen in the dual space V^* , we call the mapping $u \rightarrow u^*$ the *duality mapping*. This mapping is norm-preserving but is generally non-linear. The pairing of two vector spaces V and V^* is to specify a particular $T = u^*$ such that $T(u) = u^*(u) = \gamma(\|u\|_V)\|u\|_V$. The duality mapping may be linear when the gauge function $\gamma(t) = ct$ is linear and when the norm $\|\cdot\|_V$ is a special kind – this is the case of Hilbert space to be discussed next.

8.2.1.2 Inner Products and Hilbert Spaces

Hilbert spaces are a special case of normed vector spaces that are well-understood and widely used in physics and engineering applications. A Hilbert space is characterized by the existence of an inner product operation on a vector space. The inner product can uniquely induce a norm, a Hilbert 2-norm; conversely, a norm satisfying a certain equality (see below) can uniquely induce an inner product. Equivalently, a Hilbert space is a vector space that is self-dual, with linear duality mapping. The relationship between the existence of an inner product and self-dual property is reviewed next.

An inner product on a vector space V , denoted by $\langle \cdot, \cdot \rangle_V$ or simply $\langle \cdot, \cdot \rangle$, is by definition a function mapping $V \times V$ to \mathbb{C} such that for all $u, v, w \in V$ and $c \in \mathbb{C}$

- (i) *non-negativity*: $\langle u, u \rangle \geq 0$ for all $u \in V$ and $\langle u, u \rangle = 0$ if and only if $u = 0$;
- (ii) *linearity about the first variable*: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle cu, v \rangle = c\langle u, v \rangle$;
- (iii) *conjugate symmetry*: $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

It can be shown that (ii) and (iii) lead to linearity about the second variable in $\langle u, v \rangle$.

An inner product $\langle \cdot, \cdot \rangle$ on a vector space V induces a norm on V by

$$\|u\| := \langle u, u \rangle^{1/2}, \quad u \in V.$$

One can prove that the induced quantity is indeed a norm, that is, it satisfies the three axioms for norms listed in Section 8.2.1.1. Furthermore, this norm *as induced from the inner product* satisfies the “parallelogram law” (also known as “polarization identity”)

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2), \quad u, v \in V.$$

Conversely, a norm on V induces an inner product:

$$\langle u, v \rangle = \frac{1}{4}(\|u + v\|^2 - \|u - v\|^2 + i\|u + iv\|^2 - i\|u - iv\|^2)$$

when the norm $\|\cdot\|$ satisfies the parallelogram law. A vector space equipped with an inner product is called an *inner product space*. An inner product space whose induced norm is complete is called a Hilbert space.

On an inner product space V over \mathbb{R} , the angle between two non-zero vectors u, v can be measured by

$$\arccos \frac{\langle u, v \rangle}{\|u\| \|v\|}.$$

Thus, two vectors $u, v \in V$ are said to be orthogonal to each other if $\langle u, v \rangle = 0$. This leads to the notion of “orthogonal projection” of a vector u onto a closed subspace V_0 – it is the vector $v_0 \in V_0$ such that their difference $u - v_0$ is orthogonal to each vector in V_0 :

$$\langle u - v_0, v \rangle = 0, \quad \forall v \in V_0.$$

The orthogonal projection v_0 of u coincides with the best approximation of u in V_0 , that is,

$$\|u - v_0\| = \inf_{v \in V_0} \|u - v\|.$$

Note that, fixing one slot of $\langle \cdot, \cdot \rangle$ turns the inner product into a linear functional $T(\cdot) = \langle \cdot, u_0 \rangle$ upon fixing u_0 – a linear functional can be viewed as an u_0 -indexed inner product for some vector $u_0 \in V$. Furthermore, the dependency on u_0 is linear, due to linearity with respect to the second variable in an inner product – there is a linear duality map between $T \in V^*$ and $u_0 \in V$. It turns out that every continuous linear functional arises this way, as guaranteed by the fundamental theorem in functional analysis, namely the Riesz representation theorem:

Theorem 8.1 (Riesz representation theorem for Hilbert spaces) *Every continuous linear functional v^* on a Hilbert space V is representable as a v -indexed inner product $v^*(\cdot) = \langle \cdot, v \rangle$. In other words, every element in the dual space V^* of linear functionals on V can be represented by an element of the original vector space V .*

To summarize, a Hilbert space can be equivalently defined as a normed vector space

- (i) equipped with an inner product; or
- (ii) with norm-satisfying parallelogram law; or
- (iii) with duality mapping being linear.

8.2.1.3 Semi-inner Products and Banach Spaces

Moving beyond Hilbert spaces means that we have to sacrifice “inner product” operation and all the nice (linear) properties it brings. This is the setting of Banach spaces, which deal exclusively with “norm” (which is a non-linear functional) and norm-induced metric. More precisely, Banach spaces are normed vector spaces in which the norm-induced metric is “complete” (we do not define “complete” here, since it deals with the issue of denseness of subspaces). It suffices to think of a Banach space to be well-behaved both as a vector space and as a metric space.

A Hilbert space is a Banach space with an inner product. So the only new operation on a vector space V that is introduced into a Hilbert space is an inner product. Hence, in studying general Banach spaces, we need to carefully delineate which properties of a Hilbert space are due to the existence of an inner product, and which are generic properties of a Banach space. An important bridge is provided by the less-popular notion of “semi-inner product” on a normed vector space.

Semi-inner products were introduced into mathematics (Giles, 1967; Lumer, 1961) for the purpose of extending the geometric structure of inner product spaces to general vector spaces. A semi-inner product on a vector space V , hereby denoted by $[\cdot, \cdot]$, is a mapping from $V \times V$ to \mathbb{C} such that for all $u, v, w \in V$ and $c \in \mathbb{C}$

- (i) *non-negativity*: $\langle u, u \rangle \geq 0$ for all $u \in V$ and $\langle u, u \rangle = 0$ if and only if $u = 0$;
- (ii) *linearity about the first variable*: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle cu, v \rangle = c\langle u, v \rangle$;
- (iii) *Cauchy–Schwarz inequality*: $|[u, v]|^2 \leq [u, u][v, v]$.

Compared with the definition of an inner product $\langle \cdot, \cdot \rangle$, the first two stipulations are identical, and only (iii) differs: the only property of inner products that a semi-inner product need not possess is conjugate symmetry, that is, $[u, v] \neq \overline{[v, u]}$, which is equivalent to non-linearity with respect to the second argument: $[w, u + v] \neq [w, u] + [w, v]$. Quite pleasingly, $[u, u]^{1/2} \equiv \|u\|$ defines a norm on V . Furthermore, as Giles (1967) showed, $[u, cv] = \bar{c}[u, v]$, that is, the semi-inner product is (conjugate) homogeneous with respect to its second argument. As an example, we give here semi-inner product on the space of all sequences $\ell^1(\mathbb{N})$ under ℓ^1 -norm:

$$[u, v] := \|v\|_{\ell^1} \sum_{j \in \mathbb{N}} u_j \operatorname{sgn}(v_j), \quad u, v \in \ell^1(\mathbb{N})$$

where $\operatorname{sgn}(t)$ denotes the sign of $t \in \mathbb{R}$.

A semi-inner product on a vector space is a surrogate to the inner product. It plays similar roles as an inner product by providing a representation for continuous linear functionals (elements of the dual space) by elements of the original vector space (Riesz representation theorem). James (1964) showed an analogue to (in fact

an extension of) the classical Riesz representation theorem: every element of the dual space V^* can be “represented” by a unique element in V . Formally,

Theorem 8.2 (Riesz representation theorem for Banach space) *Assume that V is a reflexive Banach space. Every continuous linear functional v^* on V is representable as a v -indexed semi-inner product:*

$$v^*(u) = [u, v] \text{ for all } u \in V.$$

(Note that linearity with respect to the first argument of $[u, v]$ ensures that v^* is a linear functional.)

Semi-inner products also provide a means for measurement of angle (in particular, defining a kind of orthogonality) between two vectors. In the absence of an inner product, one may still have a notion of “pseudo-orthogonality” (James, 1964), extending that of orthogonality:

Definition 8.3 A vector u is said to be *pseudo-orthogonal* with respect to v in a vector space V if $\|u + cv\| \geq \|u\|$ for all $c \in \mathbb{C}$.

Giles (1967) showed that u is pseudo-orthogonal to v if and only if $[v, u] = 0$. Pseudo-orthogonality is not a symmetric relation. The best approximation property, nevertheless, is still captured: v_0 is the best approximation of u to a closed subspace V_0 in V if and only if their difference $u - v_0$ is pseudo-orthogonal to V ; the computation of v_0 involves a system of non-linear equations.

8.2.1.4 Duality Mapping and Generalized Semi-inner Products

The essence of semi-inner products is reflected in the duality mapping. Recall that starting from a vector space V , we defined its “dual” V^* , which is a vector space. One can continue this process and define the “double dual” V^{**} as the dual of V^* , two kinds of “triple duals,” respectively, $(V^{**})^*$ as the dual of V^{**} and $(V^*)^{**}$ as the double dual of V^* , etc. It turns out that first, $(V^{**})^* = (V^*)^{**}$ so the notation V^{***} can be used unambiguously, and second, $V^{***} = V^*$. Here equality is in the sense of a “linear isomorphism.” So what may appear at first as an endless process yields only two new vector spaces V^* and V^{**} , in addition to the vector space V one starts with. In general, $V \neq V^{**}$; it can be shown that V^{**} is larger than V and includes the latter as a subspace. When $V^{**} = V$, then the vector space V is called *reflexive*. When V^* and V are linearly isomorphic $V^* = V$, the vector space is called *self-dual*; this happens when the duality mapping is linear – in this case, the vector space is equipped with an inner product.

The standard semi-inner product introduced by Lumer (1961) in essence provides a “canonical” duality map between V and V^* , i.e., canonical in the sense of satisfying $\|v^*\|_{V^*} = \|v\|_V$. Nath (1971) relaxed the Cauchy–Schwarz inequality to the Hölder inequality and obtained a definition of semi-inner products of order p :

(iii) *Hölder inequality:* $|[u, v]| \leq ([u, u])^{1/p} ([v, v])^{1/q}, \quad u, v \in V,$

for $p, q \in (1, \infty)$ and satisfying $p^{-1} + q^{-1} = 1$. Nath's definition amounts to relaxing the duality mapping to $\|v^*\|_{V^*} = \|v\|_V^{p-1}$.

Zhang and Zhang (2010) investigated the most general form of a semi-inner product that might be compatible with a norm. Here, compatibility means that $[v, v]$ is indeed monotonically related to the norm $\|v\|_V$ endowed on V . It turns out that Condition (iii) can be extended to the following inequality:

(iii) *generalization of Hölder inequality*: $[[u, v]] \leq \varphi([u, u])\psi([v, v]), \quad u, v \in V,$

where φ and ψ satisfy $\varphi(t)\psi(t) = t$, and some other conditions, including monotonicity. This corresponds to a generalized duality mapping requiring $\|v^*\|_{V^*} = \varphi^{-1}(\|v\|_V)/\|v\|_V$.

8.2.1.5 Bases in Hilbert and Banach Spaces

In a vector space V , one can find a set $E = \{u_j : j \in \mathbb{I}\}$ of linearly independent vectors to serve as its basis, so that each vector $v \in V$ can be expressed as a sequence of numbers or "coordinates" using the basis (here we assume countable basis, with $\mathbb{I} = \{1, 2, \dots, m\}$ an index set allowing $m = \infty$). Linear independence is in the sense that (i) no element (vector) in E can be written as a linear combination of other elements in it; and (ii) all elements of V can be written as a unique linear combination of elements of E .

When V is a Hilbert space with an inner product $\langle \cdot, \cdot \rangle$, it turns out that one can always construct an orthonormal basis such that: $\langle u_i, u_j \rangle = \delta_{ij}$, the Kronecker delta. In this case each vector in V has a simple decomposition:

$$u = \sum_{j \in \mathbb{I}} \langle u, u_j \rangle u_j.$$

In other words, the coordinates are simply given by projection of the vector u onto each of the basis vectors u_j 's.

For a general vector space (not being equipped with an inner product), due to the lack of definition of orthogonality, we cannot hope for the existence of such an orthonormal basis. Instead, when V is reflexive, we can use the semi-inner product to construct a pair of pseudo-orthogonal bases $\{u_j : j \in \mathbb{I}\}$ and $\{v_j : j \in \mathbb{I}\}$:

$$[u_j, v_k] = \delta_{jk},$$

where pseudo-orthogonality between two vectors is in the sense of James discussed earlier. In analogy to the Hilbert space situation with an inner product, any element $u \in V$ of a Banach space with a semi-inner product can be decomposed as

$$u = \sum_{j \in \mathbb{I}} [u, v_j] u_j.$$

Note that the projection ($c_j = [u, v_j]$) and reconstruction ($u = \sum_j c_j u_j$) use two different sets of vectors (as opposed to one set, in the case of the Hilbert base).

8.2.1.6 Frames in Hilbert and Banach Spaces

A basis of a vector space V is a minimum set that can be used to represent or reconstruct any element of V . The vectors in a basis “span” the entire space V . This is the irredundant representation of vectors, which is unique. Sometimes, it is convenient to consider over-redundant (and hence non-unique) representations, by allowing more vectors in the set than an independent set. Such representation is called a “frame,” which brings more flexibility in representing and approximating vectors.

Formally, a collection of vectors $\{v_j \in V : j \in \mathbb{I}\}$ is called a frame for a Hilbert space V if there exist two constants α, β (with $0 < \alpha \leq \beta$) such that

$$\alpha \|u\|^2 \leq \sum_{j \in \mathbb{I}} |\langle u, v_j \rangle|^2 \leq \beta \|u\|^2$$

for all $u \in V$. When the frame $\{v_j : j \in \mathbb{I}\}$ satisfies linear independence, i.e., none of its members can be expressed as a linear combination of other members

$$v_j \notin \overline{\text{span}}\{v_k : k \in \mathbb{I}; k \neq j\},$$

the frame is called a Riesz basis for a Hilbert vector space V .

Given a frame in a Hilbert space, we can construct an “analysis operator” $\mathcal{A} : V \rightarrow \ell^2(\mathbb{I})$ defined by

$$\mathcal{A}u := \{\langle u, v_j \rangle : j \in \mathbb{I}\}.$$

and a “synthesis operator” from $\ell^2(\mathbb{I})$ to V , as the dual operator \mathcal{A}^\dagger of \mathcal{A} , given by

$$\mathcal{A}^\dagger c := \sum_{j \in \mathbb{I}} c_j v_j.$$

By the definition of frames, the operator $S := \mathcal{A}^\dagger \mathcal{A}$ satisfies, for $\forall u \in V$,

$$\alpha \|u\|^2 \leq \langle Su, u \rangle \leq \beta \|u\|^2.$$

Thus, S is positive and has a bounded inverse. It is used to provide a sampling formula

$$u = S^{-1}Su = \sum_{j \in \mathbb{I}} \langle u, v_j \rangle (S^{-1}v_j), \quad u \in V. \quad (8.1)$$

In a Hilbert space V , a dual frame of $\{v_j : j \in \mathbb{I}\}$ is any frame $\{u_j : j \in \mathbb{I}\}$ for V such that for all $u \in \mathcal{H}$

$$u = \sum_{j \in \mathbb{I}} \langle u, v_j \rangle u_j.$$

By Equation (8.1), $\{S^{-1}v_j : j \in \mathbb{I}\}$ is a dual frame of $\{v_j : j \in \mathbb{I}\}$. It is called the canonical dual frame of $\{v_j : j \in \mathbb{I}\}$, where “canonical dual” is in the L^2 sense (Li, 1995). In the case of Banach spaces, frames and Riesz bases can still be defined analogously despite of a lack of the inner product operator, which is instrumental for defining a Hilbert space frame and Riesz basis. Zhang and Zhang (2011) provided a definition using the semi-inner product, improving an earlier characterization by Casazza, Christensen, and Stoeva (2005).

Definition 8.4 Let X_d be a chosen discrete sequence space. We call $\{v_j : j \in \mathbb{I}\} \subseteq V$ a *frame* for V modeled on X_d if $\{[u, v_j] : j \in \mathbb{I}\} \in X_d$ for all $u \in V$ and there exist two constants α, β (with $0 < \alpha \leq \beta$) such that

$$\alpha \|u\|_V \leq \|\{[u, v_j] : j \in \mathbb{I}\}\|_{X_d} \leq \beta \|u\|_V \quad \text{for all } u \in V.$$

We call $\{v_j : j \in \mathbb{I}\} \subseteq V$ a *Riesz basis* for V modeled on X_d if $\text{span}\{v_j\} = V$, $\sum_j c_j v_j$ converges in X_d for all $c \in X_d$, and there exist two constants α, β (with $0 < \alpha \leq \beta$) such that

$$\alpha \|c\|_{X_d} \leq \left\| \sum_{j \in \mathbb{I}} c_j v_j \right\|_V \leq \beta \|c\|_{X_d} \quad \text{for all } c \in X_d.$$

Analogous expressions for analysis operator and synthesis operator were established for V and V^* , generalizing those for Hilbert spaces (Zhang & Zhang, 2011).

8.2.2 Function Spaces and Reproducing Kernels

Having introduced vector spaces and studied their properties, we now focus on a special kind of vector space, called “function spaces,” in which elements are functions taking inputs from an arbitrary set. More specifically, we study the set $\mathbb{C}^{\mathcal{X}}$ of all complex-valued functions on a prescribed set \mathcal{X} . That the set of all functions $f : \mathcal{X} \rightarrow \mathbb{C}$ forms a vector space in its own right can be seen by verifying all the requirements of a vector space, where vector addition (i) and scalar multiplication (ii) are, respectively, defined by

$$(f + g)(x) = f(x) + g(x), \quad x \in \mathcal{X}$$

and

$$(cf)(x) = c(f(x)), \quad x \in \mathcal{X}$$

for $f, g \in \mathbb{C}^{\mathcal{X}}, c \in \mathbb{C}$. Below, these functions are also referred to as \mathcal{X} -functions, to remind the reader of the fact that the two operations (i) and (ii) are defined “point-wise” on \mathcal{X} in forming the vector space \mathcal{B} of functions,

$$\mathcal{B} = \{f : f \text{ is a } \mathcal{X}\text{-function}\},$$

regardless of the structure of the prescribed input space \mathcal{X} .

Recall (from Section 8.2.1.1) that a functional $T : V \rightarrow \mathbb{C}$ takes its inputs from a vector space V (whereas in general, a function $f : \mathcal{X} \rightarrow \mathbb{C}$ takes its inputs from an arbitrary space \mathcal{X}). In the case when the underlying vector space (serving as the input space for functionals) is a space of functions, then a functional maps any function to a number; a functional can be thought of as a “function of a function” (as reflected by the suffix “-al”). For example, with respect to the function space \mathcal{B} (on some prescribed set \mathcal{X} , without any structural assumptions about \mathcal{X}), a functional (whether linear or non-linear) on \mathcal{B} takes any \mathcal{X} -function $f \in \mathcal{B}$ and produces a number.

Norm $\|\cdot\|_{\mathcal{B}}$ of the function space \mathcal{B} is an example of non-linear functionals, following the axioms that define the norm $\|\cdot\|_V$ of any vector space V . An important example of linear functionals is the “point evaluation” of a function. A point evaluation functional $T_{\text{ev}(\cdot)} : \mathcal{B} \rightarrow \mathbb{C}$ takes as input a \mathcal{X} -function $f : \mathcal{X} \rightarrow \mathbb{C}$ (an element of the function space \mathcal{B}) and gives as output a number (an element of \mathbb{C}) by restricting the \mathcal{X} -function at an evaluated point x_0 :

$$T_{\text{ev}(x_0)}(f) = f(x)|_{x=x_0} = f(x_0)$$

where $x_0 \in \mathcal{X}$ is an element of the prescribed input space \mathcal{X} . The point evaluation functional is, of course, indexed by the given point x_0 . The notion of “continuity” for evaluation functionals relies on the norm-induced topology on \mathcal{B} as a vector space.

Apart from evaluation functionals, function spaces may also have “reproducing kernels.” For a Banach space \mathcal{B} (including Hilbert space \mathcal{H}) of functions $\mathcal{X} \rightarrow \mathbb{C}$, a reproducing kernel K (“kernel” for short) is a bivariate function: $\mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C}$ that satisfies additional conditions. A consequence of Riesz representation theorem (Theorem 8.1) (and its Banach space extension, Theorem 8.2) is that point evaluation can be represented by an inner product (and its Banach space extension, semi-inner product) through the use of reproducing kernels. This is the mathematical foundation of theories of sampling and approximation, and the starting point for developing kernel methods for regularized learning, whose goal is to generalize point evaluation from given sample inputs to all points (in a prescribed input space \mathcal{X}).

8.2.2.1 Reproducing Kernel Hilbert Spaces (RKHS)

In a Hilbert space \mathcal{H} of functions, reproducing kernels are such that $K(x_0, \cdot) \in \mathcal{H}$, $K(\cdot, x_0) \in \mathcal{H}$ for all $x_0 \in \mathcal{X}$. Reproducing kernels offer a convenient device for “sampling” the function at input points. The reproducing property of the kernel K links any function $f : \mathcal{X} \rightarrow \mathbb{C}$ to its point evaluation:

$$f(x_0) = \langle f(\cdot), K(\cdot, x_0) \rangle = \langle K(x_0, \cdot), f(\cdot) \rangle \quad (8.2)$$

with $\langle \cdot, \cdot \rangle$ denoting the inner product of the RKHS satisfying $K(x, x') = K(x', x)$ and

$$K(x, x') = \langle K(x, \cdot), K(\cdot, x') \rangle.$$

For this reason, the reproducing kernel plays the role of an “evaluation functional” – given a function and a point $x_0 \in \mathcal{X}$ of the sample space, K is used to return a number which is the function’s value $f(x_0)$ at that point. A reproducing kernel Hilbert space (RKHS) is defined as a function space in which point evaluations are continuous. Every RKHS is completely characterized by a kernel function K ; the space is normed by $\|\cdot\|_{\mathcal{H}_K}$, a 2-norm as induced by its inner product $\langle \cdot, \cdot \rangle$:

$$K(x, x) = \langle K(x, \cdot), K(\cdot, x) \rangle = \|K(x, \cdot)\|_{\mathcal{H}_K}^2.$$

(Note the 2-norm is on the Hilbert space of functions $\mathcal{X} \rightarrow \mathcal{C}$, and not on the input space \mathcal{X} .) Well-known examples of kernel functions are polynomials, radial basis functions, etc. It is long known that a bivariate function K can become a reproducing kernel if and only if (i) the corresponding kernel matrix $K(x_i, x_j)$, $i = 1, \dots, n$ is positive semi-definite for any arbitrary set of sample points in \mathcal{X} and for any set-size n (Aronszajn, 1950), or that (ii) there exists a mapping Φ from \mathcal{X} to some inner product space \mathcal{W} such that

$$K(x, x') = \langle \Phi(x), \Phi(x') \rangle_{\mathcal{W}}, \quad x, x' \in \mathcal{X}.$$

8.2.2.2 Reproducing Kernel Banach Spaces (RKBS)

We define an RKBS as a reflexive Banach space \mathcal{B} of functions on \mathcal{X} such that the dual space \mathcal{B}^* is isometrically isomorphic to (and hence can be identified with) a Banach space of functions on \mathcal{X} and point evaluations are continuous linear functionals on both \mathcal{B} and \mathcal{B}^* (as a space of functions). A bivariate function $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{C}$ is a reproducing kernel in an RKBS if and only if there exist mappings Φ from \mathcal{X} to some reflexive Banach space \mathcal{W} and $\Phi^* : \mathcal{X} \rightarrow \mathcal{W}^*$ such that

$$K(x, x') = \langle \Phi(x), \Phi^*(x') \rangle_{\mathcal{W}}, \quad x, x' \in \mathcal{X},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ denotes the “dual pairing” of \mathcal{W} with \mathcal{W}^* , which is a bilinear form. In this case, $K(\cdot, \cdot) \in \mathcal{B}$, $K(\cdot, x) \in \mathcal{B}^*$ for all $x \in \mathcal{X}$, and K has the following reproducing properties:

$$f(x) = (f, K(\cdot, x))_{\mathcal{B}}, \quad f^*(x) = (K(x, \cdot), f^*)_{\mathcal{B}}, \quad f \in \mathcal{B}, \quad f^* \in \mathcal{B}^*.$$

Here $\langle \cdot, \cdot \rangle_{\mathcal{B}}$ is a bilinear form (“dual pairing”) on $\mathcal{B} \times \mathcal{B}^*$:

$$(f, g^*) := g^*(f), \quad f \in \mathcal{B}, \quad g^* \in \mathcal{B}^*.$$

As a special case of RKBS, we have the reproducing kernel Hilbert space (RKHS), where \mathcal{B} is simply a Hilbert space \mathcal{H} of functions. A Hilbert space is a Banach space with the addition of *any* of the following conditions:

- (i) the semi-inner product $[\cdot, \cdot]$ on \mathcal{B} is conjugate symmetric: $[f, g] = \overline{[g, f]}$ (and hence is an inner product);
- (ii) the dual space \mathcal{B}^* of \mathcal{B} is isometrically isomorphic to \mathcal{B} ;
- (iii) the duality mapping from $\mathcal{B} \ni f \mapsto f^* \in \mathcal{B}^*$ is linear.

8.2.2.3 Operator-valued Reproducing Kernels

In the above discussions, the Banach space \mathcal{B} of functions f is scalar-valued, $f : \mathcal{X} \rightarrow \mathbb{C}$. In machine learning applications (especially kernel learning), there is a need to consider vector-valued functions $f : \mathcal{X} \rightarrow \mathbb{C}^m$, where m is the dimension of the output space $\Lambda = \mathbb{C}^m$.

The first step is to extend the reproducing kernel Hilbert space (RKHS) to a vector-valued RKHS, as formally defined below (Micchelli & Pontil, 2005a,b).

Definition 8.5 A Λ -valued RKHS is a Hilbert space \mathcal{H} of functions from the input space \mathcal{X} to the output vector space Λ such that for each $x \in \mathcal{X}$ and for each $f \in \mathcal{H}$, the point evaluation $f \mapsto f(x)$ is a continuous linear operator from \mathcal{H} to Λ .

In this definition, \mathcal{H} is the space of vector-valued functions with an inner product operation $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ taking any two such functions into a number \mathbb{C} . A kernel \mathcal{K} in a Λ -valued RKHS is a bivariate, operator-valued function of \mathcal{X} such that, when evaluated in each variable, the operator-valued functions $\mathcal{K}(\cdot, x)$ and $\mathcal{K}(x, \cdot)$ satisfy $\forall u \in \Lambda$

$$\langle f(x), u \rangle_{\Lambda} = \langle f(\cdot), \mathcal{K}(\cdot, x)u \rangle_{\mathcal{H}}, \quad \langle \mathcal{K}(x, \cdot)u, f(\cdot) \rangle_{\mathcal{H}} = \langle u, f(x) \rangle_{\Lambda}, \quad (8.3)$$

and hence they are equal. Here, $\mathcal{K}(x, x')$ is a linear operator (an $m \times m$ matrix) from Λ to Λ ; when we speak of $\mathcal{K}(\cdot, \cdot)$ as a bivariate function, we mean that it takes input from $\mathcal{X} \times \mathcal{X}$ and outputs such an operator over vector space Λ , that is, an $m \times m$ matrix. \mathcal{K} serves as a point evaluation map for the vector-valued function f . The kernel \mathcal{K} satisfies the property that $\forall u, v \in \Lambda$

$$\langle u, \mathcal{K}(x, x')v \rangle_{\Lambda} = \langle \mathcal{K}(x, \cdot)u, \mathcal{K}(\cdot, x')v \rangle_{\mathcal{H}} = \langle \mathcal{K}(x', x)u, v \rangle_{\Lambda} \quad (8.4)$$

and that $\mathcal{K}(x, x') = (\mathcal{K}(x', x))^{\dagger}$, where \dagger stands for adjoint transformation of an operator (i.e., conjugate transpose of a matrix) in the sense that

$$\langle \mathcal{K}u, v \rangle_{\Lambda} = \langle u, \mathcal{K}^{\dagger}v \rangle_{\Lambda}. \quad (8.5)$$

Clearly, $\mathcal{K}(x, x)$ is self-adjoint and hence non-negative, satisfying the following:

$$|\langle \mathcal{K}(x, x')u, v \rangle_{\Lambda}| \leq \langle \mathcal{K}(x, x)u, u \rangle_{\Lambda}^{1/2} \langle \mathcal{K}(x', x')v, v \rangle_{\Lambda}^{1/2} = \|\mathcal{K}(x, \cdot)u\|_{\mathcal{H}} \|\mathcal{K}(x', \cdot)v\|_{\mathcal{H}}. \quad (8.6)$$

Micchelli and Pontil (2005a,b) showed that an operator-valued bivariate function \mathcal{K} is a reproducing kernel if and only if for any sequence ($i \in \mathbb{I}$) of points $x_i \in \mathcal{X}$ and any sequence of bounded vectors $u_i \in \Lambda$,

$$\sum_{i,j} \langle \mathcal{K}(x_i, x_j)u_j, u_i \rangle_{\Lambda} \geq 0.$$

This is analogous to Aronszajn's (1950) characterization of reproducing kernels for scalar-valued Hilbert spaces.

The above result has been extended to vector-valued reproducing kernel Banach spaces in Zhang and Zhang (2013), with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ in Equations (8.3), (8.4) and (8.6) replaced by semi-inner product $[\cdot, \cdot]_{\mathcal{B}}$. In particular, instead of Equation (8.5), the *generalized adjoint* \mathcal{K}^{\dagger} of an operator \mathcal{K} is defined by

$$[\mathcal{K}u, v]_{\Lambda} = [u, \mathcal{K}^{\dagger}v]_{\Lambda}. \quad (8.7)$$

8.2.2.4 Measurable Functions as Vector Spaces: $L^p(\mathcal{X})$, $M(\mathcal{X})$, $C_0(\mathcal{X})$, etc.

When the underlying set \mathcal{X} is endowed with a measure, the space of measurable functions on \mathcal{X} has further properties. For example, the space $L^p(\mathcal{X}, \Omega, \mu)$ ($1 \leq p < +\infty$) of μ -measurable functions on \mathcal{X} , when endowed with the norm

$$\|f\|_{L^p} := \left(\int_{\Omega} |f(x)|^p d\mu \right)^{1/p} < +\infty,$$

is a normed vector space. When $p = +\infty$, $L^\infty(\mathcal{X}, \Omega, \mu)$ can be endowed with the following norm

$$\|f\|_{L^\infty} = \inf\{\alpha > 0 : \{x : |f(x)| > \alpha\} \text{ is a set with measure zero with respect to } \mu\}.$$

Generalizations of the above L^p spaces include the Orlicz spaces (Rao & Ren, 1991) and L^p spaces over other measure spaces.

The dual space of $L^p(\mathcal{X}, \Omega, \mu)$ when $1 < p < +\infty$ can be identified with $L^q(\mathcal{X}, \Omega, \mu)$, where q is the conjugate number of p satisfying $p^{-1} + q^{-1} = 1$. More accurately, by identification, we mean that there exists a mapping $L^q \ni g \mapsto \int_{\Omega} f g d\mu$, indexed by f , that defines an isometric isomorphism of $L^q(\mathcal{X}, \Omega, \mu)$ onto $(L^p(\mathcal{X}, \Omega, \mu))^*$. The duality mapping from an element f in $L^p(\mathcal{X}, \Omega, \mu)$ ($1 < p < +\infty$) to an element f^* of $L^q(\mathcal{X}, \Omega, \mu)$ is given by

$$f \rightarrow f^* = \frac{\bar{f}|f|^{p-2}}{\|f\|_{L^p}^{p-2}}.$$

Semi-inner products on $L^p(\mathcal{X}, \Omega, \mu)$, ($1 \leq p < \infty$) take the form:

$$[g, f] := \frac{\int_{\Omega} g \bar{f} |f|^{p-2} d\mu}{\|f\|_{L^p}^{p-2}}.$$

The space $L^2(\mathcal{X}, \Omega, \mu)$, a special case of $L^p(\mathcal{X}, \Omega, \mu)$, is a Hilbert space.

Complications arise when $p = 1$ and $q = \infty$. The dual space of $L^1(\mathcal{X}, \Omega, \mu)$ is $L^\infty(\mathcal{X}, \Omega, \mu)$ when μ is σ -finite, so the latter maps onto $(L^1(\mathcal{X}, \Omega, \mu))^*$ isometrically through the map: $L^\infty \ni g \mapsto \int_{\Omega} f g d\mu$. However, neither L^1 nor L^∞ is reflexive as long as they are infinite-dimensional. As a matter of fact, $(L^\infty(\mathcal{X}, \Omega, \mu))^* = G(\mathcal{X}, \mu)$, where $G(\mathcal{X}, \mu)$ is the set of finitely additive measures on \mathcal{X} that are absolutely continuous w.r.t. μ and have finite total variation. In general, the space $G(\mathcal{X}, \mu)$ is strictly larger than $L^1(\mathcal{X}, \Omega, \mu)$.

An important subspace of $L^\infty(\mathcal{X}, \Omega, \mu)$ is the space $C_0(\mathcal{X})$ of continuous functions f on \mathcal{X} such that for all $\varepsilon > 0$, $\{x \in \mathcal{X} : |f(x)| \geq \varepsilon\}$ is compact, endowed with the sup norm

$$\|f\|_{C_0} = \sup_{x \in \mathcal{X}} |f(x)|.$$

The dual of $C_0(\mathcal{X})$ is $M(\mathcal{X})$, the set of regular Borel measures on \mathcal{X} with total variation norm (note the Borel measure can be real or complex-valued). For $\mu \in M(\mathcal{X})$, $\|\mu\|$ is defined as $|\mu|(\mathcal{X}) = \mu_+(\mathcal{X}) + \mu_-(\mathcal{X})$. The mapping $M(\mathcal{X}) \ni \mu \mapsto \int_{\mathcal{X}} f d\mu$ is an isometric isomorphism of $M(\mathcal{X})$ onto $C_0(\mathcal{X})^*$. The dual of $M(\mathcal{X})$ (as a normed vector space) is $L^\infty(M(\mathcal{X}))$, see Conway (1990, p. 76).

When \mathcal{X} is discrete, then $\ell^1 = L^1(\mathbb{N}, 2^\mathbb{N}, \mu) = M(\mathbb{N})$, whose dual space is $c_0 = C_0(\mathbb{N})$, with $c_0 \subset l^\infty = L^\infty(\mathbb{N}, 2^\mathbb{N}, \mu)$ and μ having measure 1 at each point in \mathbb{N} .

From measure theory, the specification of measurable functions depends on the measure itself. If $(\mathcal{X}, \Omega, \nu)$ is a σ -finite measure space, and μ is a complex-valued measure on (\mathcal{X}, Ω) that is absolutely continuous w.r.t. ν , then there is a unique complex-valued function f in $L^1(\mathcal{X}, \Omega, \nu)$ such that $\mu(\Delta) = \int_{\Delta} f d\nu$ for every $\Delta \in \Omega$. The quantity $f = d\mu/d\nu$ is called the Radon–Nikodym derivative. We have the following facts: (i) for any $g \in L^1(\mathcal{X}, \Omega, \mu)$ then $gf \in L^1(\mathcal{X}, \Omega, \nu)$ and $\int g d\mu = \int g f d\nu$; (ii) for $\Delta \in \Omega$, $|\mu|(\Delta) = \int_{\Delta} |f| d\nu$. This is to say, $|d\mu/d\nu| = d|\mu|/d\nu$; (iii) for $\Delta \in \Omega$, there exists an f such that $|f| = 1$ a.e. w.r.t. $|\mu|$, and $\mu(\Delta) = \int_{\Delta} f d|\mu|$.

8.2.2.5 Integral Operators and Spectral Methods

For a Hilbert space of measurable functions, a key operator that had been studied (see, e.g., Smale & Zhou, 2007) is the linear integral operator (“Mercer map”) $T_\mu : L^2(\mathcal{X}, \mu) \rightarrow L^2(\mathcal{X}, \mu)$:

$$(T_\mu f)(\cdot) \equiv \int_{\mathcal{X}} K(\cdot, y) f(y) d\mu(y) \quad (8.8)$$

associated with a positive semi-definite kernel K . When \mathcal{X} is compact, the so-called Mercer’s theorem guarantees that the spectrum of T is discrete with the existence of an orthonormal basis. In fact, the above map (8.8) can be shown to be injective, with the range of the operator T_μ exactly being the RKHS with kernel K . This is the mathematical foundation underlying the heat equation/graph Laplacian approach to manifold learning (Coifman & Maggioni, 2006). The class of so-called “spectral clustering” methods in machine learning uses graph Laplacian as the kernel (Ng, Jordan, & Weiss, 2002; Weiss, 1999).

To introduce the details of the Mercer theorem, let us assume that \mathcal{X} is a compact metric space, μ is a finite positive Borel measure on \mathcal{X} , and K is a continuous reproducing kernel on \mathcal{X} . In this case, the integral operator T_μ is bounded, compact, self-adjoint, and positive. By the spectral theorem for compact operators on a Hilbert space, there exists an orthonormal basis $\{u_j : j \in \mathbb{I}\}$ for $L^2(\mathcal{X}, \mu)$ and a sequence $\{\lambda_j : j \in \mathbb{I}\}$ of non-increasing non-negative constants that tend to zero as j goes to infinity such that

$$T_\mu u_j = \lambda_j u_j.$$

The Mercer theorem states that if μ is non-degenerate, that is, every non-empty open subset in \mathcal{X} has positive measure under μ , then

$$K(x, y) = \sum_{j \in \mathbb{N}} \lambda_j u_j(x) \overline{u_j(y)}, \quad x, y \in \mathcal{X},$$

where the series converges absolutely and uniformly for $x, y \in \mathcal{X}$.

The structure of the RKHS \mathcal{H}_K with kernel K becomes clearer in view of the above expansion. For simplicity, let us assume that all the eigenvalues λ_j of T_μ are

positive. The space \mathcal{H}_K then consists of all the functions of the form $\sum_{j \in \mathbb{N}} a_j u_j$ where

$$\sum_{j \in \mathbb{N}} \frac{|a_j|^2}{\lambda_j} < +\infty.$$

The inner product in \mathcal{H}_K takes the form

$$\left\langle \sum_{j \in \mathbb{N}} a_j u_j, \sum_{j \in \mathbb{N}} b_j u_j \right\rangle = \sum_{j \in \mathbb{N}} \frac{a_j \bar{b}_j}{\lambda_j}.$$

Considering the effect of changing measures in the Mercer theorem, we let ν be another finite positive Borel measure on \mathcal{X} that is absolutely continuous with respect to μ . Set $g = \frac{d\nu}{d\mu}$. Also denote by T_μ the integral operator associated with μ and the kernel K , and by T_ν that associated with ν and kernel

$$\tilde{K}(x, y) = \frac{K(x, y)}{\sqrt{g(x)g(y)}}, \quad x, y \in \mathcal{X}.$$

One sees that

$$T_\nu \left(\frac{u_j}{\sqrt{g}} \right) = \frac{1}{\sqrt{g}} T_\mu(u_j) = \lambda_j \frac{u_j}{\sqrt{g}} = \lambda_j \left(\frac{u_j}{\sqrt{g}} \right).$$

In other words, T_ν and T_μ have the same eigenvalues, and the eigenfunctions differ by \sqrt{g} , the square-root of the Radon–Nikodym derivative $d\nu/d\mu$.

8.3 Learning as Regularization: Problem Formulation

In this section, we rigorously formulate the problem of learning from finite samples using mathematical tools. We will review statistical learning theory, and highlight the importance of “generalization” over mere (retroactive) goodness-of-fit as the goal of learning. We emphasize decomposing errors of generalization into various sources, and investigate different regularizers used in kernel methods to achieve smoothness or sparsity objectives.

8.3.1 Statistical Learning Theory

Learning to classify inputs through finite samples is an intrinsically ill-posed problem (that is, underconstrained), and *regularization* is a key towards recovery of stable and generalizable solutions. In fact, statistical machine learning is now laid on the solid foundation of the *regularized loss minimization* framework (Cucker & Smale, 2002; Poggio & Smale, 2003). In its application to classification, the first step is to define an *empirical error* term which measures how well the model fits the data. The goal of learning is, however, not to just fit well existing data, but to generalize to new inputs. Learning theory shows that, to achieve good generalization, the empirical error should be minimized subject to (the constraint of) a data-dependent upper bound on the complexity of the function space.

8.3.1.1 Assumptions on Input and Output Spaces

The prescribed input space \mathcal{X} , such as the set of human faces, consists of a set of examples, with minimal structural assumptions. Typically, \mathcal{X} is assumed to be measurable (i.e., admits a Borel measure), so one can quantify the relative frequency that any particular example occurs. It is not necessary to assume \mathcal{X} to explicitly have a vector space structure or a metric space structure. In the reproducing kernel methods, input structure is implicitly encoded in the kernel function $K(\cdot, \cdot)$ defined on $\mathcal{X} \times \mathcal{X}$: the kernel function will have an associated vector space (i.e., feature space) and associated metric (via norm on the feature space). On the other hand, the output space \mathcal{Y} can be a number field such as \mathbb{R} (resulting in scalar-valued functions to be learnt), or a vector space \mathbb{R}^n (resulting in vector-valued functions to be learnt). The set of known pairings, i.e., samples, is denoted as $\mathcal{Z} = \{(x_j, y_j) : j = 1, \dots, n\}$. The joint probability distribution on $(\mathcal{X}, \mathcal{Y})$ is assumed to exist, and expectation with respect to it is denoted as $E_{x,y}$.

8.3.1.2 Learning as an Optimization Problem

Classification algorithms based on regularization learning solve the following optimization problem

$$\min_{f \in \mathcal{B}} \{L_{\mathcal{Z}}(f) + \lambda R(f)\} = \inf_{f \in \mathcal{B}} \frac{1}{n} \sum_{j=1}^n L(y_j, f(x_j)) + \lambda R(f) \quad (8.9)$$

where

- f denotes a learned input–output mapping, called “model”;
- \mathcal{B} is the space of possible models, i.e., the hypothesis space $\mathcal{B} \ni f$;
- \mathcal{Z} denotes the training set or sample data, consisting of known input–output pairs;
- $L_{\mathcal{Z}}(f)$ is the empirical error of the model f on the training set \mathcal{Z} ;
- R is a regularizer on the hypothesis space “penalizing” more complex models and preferring simpler ones (Occam’s razor);
- λ is a positive regularization parameter that balances the need for data-fitting and for complexity-reduction, operationalizing “optimal” generalization.

In the celebrated RKHS methods (Schölkopf & Smola, 2001; Shawe-Taylor & Cristianini, 2004), the space of models being considered consists of the Hilbert function space \mathcal{H} with a penalty of the form $R(f) = \phi(\|f\|_{\mathcal{H}})$, with the norm measuring the complexity or smoothness of the function, and $\phi(\cdot)$ a non-negative function. The sample set \mathcal{Z} consists of n pairs of points in input space \mathcal{X} and output space $\mathcal{Y} \equiv \mathbb{R}$, with empirical loss (measuring a classifier’s performance on known samples) given by $L_{\mathcal{Z}}(f) = \frac{1}{n} \sum_{j=1}^n L(f(x_j), y_j)$, $L(\cdot, \cdot) \geq 0$ being a proper loss function that achieves 0 if its two arguments take equal value.

8.3.1.3 Risk of Generalization and Error Decomposition

From a statistical learning perspective, the generalization performance of any proposed learning scheme based on finite samples is evaluated by measuring a

quantity called “excess risk.” Formally, denote $f_{\mathcal{Z},\lambda}$ as the optimal solution to Equation (8.9), we are interested in bounds of the excess risk $\Delta(f_{\mathcal{Z},\lambda})$ defined as

$$\Delta(f_{\mathcal{Z},\lambda}) = \mathbb{E}_{x,y}[L(y, f_{\mathcal{Z},\lambda}(x))] - \min_{f \in \mathcal{B}} \mathbb{E}_{x,y}[L(y, f(x))]$$

where \mathcal{B} is constructed based on the distribution $\mu(x)$ over the input space \mathcal{X} .

The excess risk $\Delta(f_{\mathcal{Z},\lambda})$ can be decomposed into three components, reflecting *sample error*, *approximation error*, and *hypothesis error* as different sources for the risk about modeling. To this end, we define the generalization error $\text{Er}(f)$ and empirical error $\text{Er}_{\mathcal{Z}}(f)$ as

$$\text{Er}(f) = \mathbb{E}_{x,y}[L(y, f(x))], \quad \text{Er}_{\mathcal{Z}}(f) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i))$$

and

$$f_* = \arg \min_{f \in \mathcal{B}} \text{Er}(f), \quad f_\lambda = \arg \min_{f \in \mathcal{B}} (\text{Er}(f) + \lambda \|f\|_{\mathcal{B}}).$$

According to Cucker and Smale (2002), the generalization error of $f_{\mathcal{Z},\lambda}$ is decomposed as

$$\Delta(f_{\mathcal{Z},\lambda}) \leq -\lambda \|f_{\mathcal{Z},\lambda}\|_{\mathcal{B}^{(n)}} + S(\mathcal{Z}, \lambda) + H(\mathcal{Z}, \lambda) + D(\lambda),$$

where sample error $S(\mathcal{Z}, \lambda)$, hypothesis error $H(\mathcal{Z}, \lambda)$, and approximation error $D(\lambda)$ are given by

$$\begin{aligned} S(\mathcal{Z}, \lambda) &= \text{Er}(f_{\mathcal{Z},\lambda}) - \text{Er}_{\mathcal{Z}}(f_{\mathcal{Z},\lambda}) + \text{Er}_{\mathcal{Z}}(f_\lambda) - \text{Er}(f_\lambda), \\ H(\mathcal{Z}, \lambda) &= \{\text{Er}_{\mathcal{Z}}(f_{\mathcal{Z},\lambda}) + \lambda \|f_{\mathcal{Z},\lambda}\|_{\mathcal{B}^{(n)}}\} - \{\text{Er}_{\mathcal{Z}}(f_\lambda) + \lambda \|f_\lambda\|_{\mathcal{B}}\}, \\ D(\lambda) &= \text{Er}(f_\lambda) - \text{Er}(f_*) + \lambda \|f_\lambda\|_{\mathcal{B}}. \end{aligned}$$

The sample error $S(\mathcal{Z}, \lambda)$ can be well bounded using the standard concentration inequality technique, and the approximation error $D(\lambda)$ is often assumed to be $O(\lambda^q)$ where $q \in (0, 1]$. The main challenge is to bound the hypothesis error $H(\mathcal{Z}, \lambda)$, which in the conventional analysis of generalization error is always upper-bounded by zero due to the fact $f_{\mathcal{Z},\lambda}$ optimizes the objective function in Equation (8.9). In Section 8.4.2, we use a sequence to model input–output mapping functions as sample size n increases. Suppressing H amounts to finding the series of Banach spaces such that $\lim_{n \rightarrow \infty} \mathcal{B}^{(n)} = \mathcal{B}$ where convergence is with respect to the probability distribution under which $\mathbb{E}_{x,y}$ is performed.

8.3.2 Regularization by Different Norms

Effective regularization is key to solving the optimal classification problem. Various norms are employed as regularizers for achieving different goals.

8.3.2.1 Norms to Enforce Smoothness

The RKHS theory provides that specifying a kernel function K amounts to specifying some feature mapping $\Phi(\cdot) : x \mapsto \Phi(x)$, where $\Phi(x)$ is an internal representation of input x , i.e., $\Phi(x)$ is a vectorial representation of x 's “features.” The kernel function itself measures the similarity between any two inputs x_i and x_j ; the measure is necessarily symmetric: $K(x_i, x_j) = K(x_j, x_i)$. Such similarity measure has the natural interpretation of projections of two vectors $\Phi(x_i)$ and $\Phi(x_j)$ onto one another in the (implicitly referenced) feature space \mathcal{W} , which is a Hilbert space itself.

There are many advantages of the RKHS approach owing to the existence of an inner product in a Hilbert space. In particular, computation of the similarity between inputs need not rely on the availability of the explicit form of Φ ; rather, similarity is typically given directly in some analytic form of a reproducing kernel K (e.g., a radial basis function). However, this L^2 -based approach has its limitations, most notably the assumed symmetry of the similarity structure (as modeled by the symmetric inner product), in violation of empirical evidence from human cognition (see below).

Besides regularization with RKHS norms, there have been other choices of norms that are Fréchet differentiable. Regularization about these norms typically results in a smooth minimization problem and thus possesses efficient numerical algorithms. Learning with such norms includes the ℓ^p -coefficient regularization with $p > 1$ (Tong, Chen, & Yang, 2010), and the smooth RKBS approach by Zhang, Xu, and Zhang (2009).

8.3.2.2 Norms to Enforce Sparsity

On par with the massive amount of work on RKHS methods, which is essentially L^2 -based, there is an equal body of work on stimulus/signal representation based on ℓ^1 -regularization and sparse approximation. Thanks to the development of Lasso in statistics (Tibshirani, 1996), compressive sensing (Candès, Romberg, & Tao, 2006), and the cubic ℓ^1 spline method for minimum norm interpolation (Lavery, 2000), it is now recognized that it is the ℓ^1 -norm that enforces a sparsity solution and optimal feature selection. Unlike RKHS, where the existence of a feature map is guaranteed but does not need to be computed, properly selecting a feature basis is crucial for the success of ℓ^1 -regularization.

It should be noted that ℓ^1 -regularization is seen as a convex surrogate to ℓ^0 -regularization, which is truly the sparsity-enforcing norm (since ℓ^0 -norm counts the number of non-zero components of a vector). Because of combinatoric explosion, ℓ^0 -regularization is computationally unfeasible, and therefore replaced by ℓ^1 -regularization (Candès & Tao, 2004).

In pursuit of further sparsity, ℓ^q -norms with $0 < q < 1$ have been considered (see, e.g., Foucart & Lai, 2009). Such semi-norms are even closer to the ℓ^0 -norm compared with the ℓ^1 -norm. In particular, $\ell^{1/2}$ -regularization (Xu *et al.*, 2010) has received considerable attention for its advantages in developing efficient numerical methods.

8.3.2.3 Different Norms on Kernelized Functions

Given arbitrary n sample points in the input space \mathcal{X} and a bivariate function K on $\mathcal{X} \times \mathcal{X}$, consider the Banach space \mathcal{B} of functions on \mathcal{X} which contains the following linear combinations of its elements

$$\sum_{j=1}^n c_j K(x_j, \cdot) \quad (8.10)$$

for arbitrary $c = \{c_j \in \mathbb{C} : j = 1, \dots, n\}$. We can endow the space \mathcal{B} with the following ℓ^1 -norm:

$$\left\| \sum_{j=1}^n c_j K(x_j, \cdot) \right\|_{\mathcal{B}} = \sum_{j=1}^n |c_j| = \|c\|_{\ell^1}, \quad (8.11)$$

or an L_2 (RKHS)-norm

$$\left\| \sum_{j=1}^n c_j K(x_j, \cdot) \right\|_{\mathcal{B}} = \left(\sum_{i,j=1}^n c_i K(x_i, x_j) c_j \right)^{1/2} = \|c\|_K. \quad (8.12)$$

Such flexibility highlights the fact that reproducing kernels are useful tools to construct the function space where functions (as vectors) may be endowed with various norms determined by practical applications.

8.4 Learning with Reproducing Kernels: Solution Concepts

This section provides an in-depth explanation of the celebrated “kernel methods.” The suite of methods include representer theorem, feature map, “kernel trick” and kernelization, maximal margin classifier, etc. These computational tools are grounded in solid mathematical techniques and derivations, and support the conceptual core for mathematical modeling of human categorization.

8.4.1 Power of Reproducing Kernels

In this section, we review the usefulness of reproducing kernels of a function space in regularized learning. Among these are (i) representer theorem, which allows the optimal solution of the regularized learning problem to be representable through the kernel function; (ii) feature map, which allows an input stimulus to have feature representation in a vector space; and (iii) kernel trick, which allows the computation to bypass exact feature representation.

8.4.1.1 Representer Theorem

The now-popular RKHS theory guarantees that the optimal classifier f_{opt} is a weighted sum of similarities (as measured through a symmetric, positive semi-definite kernel function K) to the sampled points:

$$f_{opt}(\cdot) = \sum_{j=1}^n c_j K(x_j, \cdot). \quad (8.13)$$

The above statement is known as the “representer theorem.”

In the case of RKBS in general, denote $G(\cdot, x) = K(x, \cdot)^*$. Following the discussions of Section 8.3.1.2, we formulate the following regularized risk minimization problem (under a general regularization function ϕ)

$$\inf_{f \in \mathcal{B}} \frac{1}{n} \sum_{j=1}^n L(f(x_j), y_j) + \lambda \phi(\|f\|_{\mathcal{B}}), \quad (8.14)$$

where \mathcal{B} is a Banach space of functions $f : \mathcal{X} \rightarrow \mathbb{C}$. Its solution f_{opt} is characterized by the following “representer theorem”:

Theorem 8.6 (Zhang & Zhang, 2012) *Let L be continuous and proper (i.e., $L(s, t) \geq 0$, achieving 0 if and only if $t = s$) and let ϕ be continuous and non-decreasing with $\lim_{t \rightarrow \infty} \phi(t) = +\infty$. Then (8.14) has a minimizer f_{opt} in \mathcal{B} , which has the form*

$$f_{opt}^*(\cdot) = \sum_{j=1}^n c_j K(x_j, \cdot)^* = \sum_{j=1}^n c_j G(\cdot, x_j) \quad (8.15)$$

for some constants $c_j \in \mathbb{C}, j = 1, \dots, n$. Here f_{opt}^* represents the dual of f_{opt} .

Note that, unlike the RKHS case, the kernels are generally not positive definite once the number of sampling points exceeds two. The essence of a representer theorem is to represent the dual function of the minimizer as a linear combination of the point evaluation functionals at x_j , $1 \leq j \leq n$. The fact that the representer theorem is really a statement about an element of the dual space has not been previously appreciated due to the reason that in an RKHS, $f^* = f$.

8.4.1.2 Feature Spaces and Feature Maps

If stimuli are represented as individual points in the sample space \mathcal{X} , similarity between inputs in \mathcal{X} can be measured by the inner product of their respective features in some feature space, which is a vector space denoted as \mathcal{W} . In the standard RKHS theory, the feature space \mathcal{W} is further assumed to be endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{W}}$. The feature map $\Phi(\cdot) : \mathcal{X} \rightarrow \mathcal{W}$ is captured by the reproducing kernel K :

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathcal{W}}, \quad (8.16)$$

that is necessarily symmetric: $K(x_i, x_j) = K(x_j, x_i)$. The important advantage for the RKHS framework is its *implicit* reference to a feature representation (i.e., mapping Φ) – note that inputs in \mathcal{X} for classification typically do not have a Hilbert space or even vector space structure for that matter; however, one is able to assume the existence of a (usually much higher-dimensional) feature space (which is a Hilbert space) that the input space can be mapped into.

These same considerations can be extended to Banach spaces in general. We have proven:

Theorem 8.7 (Zhang, Xu, & Zhang, 2009) *Let \mathcal{B} be a uniform Banach space of functions on an input space \mathcal{X} where point evaluations are continuous linear functionals. Then*

1. *There exists a unique function (called “semi-inner product reproducing kernel”) $G : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ such that $G(x, \cdot) \in \mathcal{B}$ for all $x \in \mathcal{X}$ and*

$$f(x) = [f, G(x, \cdot)]_{\mathcal{B}} \text{ for all } f \in \mathcal{B} \text{ and } x \in \mathcal{X}. \quad (8.17)$$

2. *A function $G : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is the reproducing kernel of \mathcal{B} if and only if there exists some mapping Φ from \mathcal{X} to a uniform Banach space \mathcal{W} such that*

$$G(x, y) = [\Phi(x), \Phi(y)]_{\mathcal{W}} \text{ for all } x, y \in \mathcal{X}. \quad (8.18)$$

Here, “uniform” means that the Banach space of functions under consideration is uniform convex and uniform Fréchet differentiable, conditions that guarantee the existence and uniqueness of a semi-inner product. Since a semi-inner product is generally non-symmetric unless the Banach space becomes a Hilbert space, similarity is non-symmetric in general – there may be x and y such that

$$G(x, y) = [\Phi(x), \Phi(y)]_{\mathcal{W}} \neq [\Phi(y), \Phi(x)]_{\mathcal{W}} = G(y, x).$$

In the case of vector-valued RKHS \mathcal{H} or vector-valued RHBS \mathcal{B} where the function under question is vector-valued (e.g., taking values in $\Lambda = \mathbb{C}^m$), the characterization of kernel \mathcal{K} is through the existence of an operator-valued mapping $\mathcal{W}(\cdot)$ from \mathcal{X} to the space of operators from the feature space \mathcal{W} to Λ (that is, the space of $m \times \dim \mathcal{W}$ matrices):

$$\mathcal{K}(x, x') = \mathcal{W}(x') \mathcal{W}^\dagger(x), \quad x, x' \in \mathcal{X}. \quad (8.19)$$

Note that \mathcal{W} is an operator-valued feature map, that is, an $m \times \dim \mathcal{W}$ matrix for each input value x . So this is the analogue to “feature map” Φ in the scalar-valued function learning, except now the “features” are operators (matrices) instead of “vectors.” When $m = 1$, \mathcal{W} is simply the map Φ ; viewed in this way, \mathcal{W} is simply a multitude of feature maps. The above representation formula says that the kernel operator $\mathcal{K}(x, x')$ is simply the multiplication of two matrices $\mathcal{W}(x')$ and $\mathcal{W}^\dagger(x)$. Micchelli and Pontil (2005a,b) derived this result for the vector-valued RKHS case, and Zhang and Zhang (2013) extended Equation (8.19) to the case of vector-valued RKBS. The only difference between the two cases is that \dagger is a generalized adjoint defined in Equation (8.7) through the use of a semi-inner product on \mathcal{B} .

8.4.1.3 Kernel Trick

Measurement of similarity of two input points, x_i and x_j , via their feature representations, $\Phi(x_i)$ and $\Phi(x_j)$, is through their inner product, or semi-inner product in general. The “kernel trick” refers to the fact that one does not need explicit

knowledge of feature representation, that is, the function form of $\Phi(\cdot)$, to compute the similarity. Rather, knowledge of the given kernel function $K(\cdot, \cdot)$ suffices. Hence, existence of the feature representation is guaranteed but not invoked in computation. It is well-known in the RKHS theory that a bivariate function makes a reproducing kernel if and only if it is representable as a similarity measurement by Equation (8.16), see Aronszajn (1950).

It should be mentioned that the kernel trick allows us to “kernelize” an input representation, so that the input space need not be a Hilbert space of vectors, but rather an arbitrary space with a kernel function. This leads to popular algorithms such as kernel principle component analysis (kPCA), kernel density estimation, kernel regression, etc. In those algorithms, what matters is the result of projection of feature vectors $\Phi(x_i)$ and $\Phi(x_j)$ onto one another, so one can use the evaluation of the kernel function $K(x_i, x_j)$ instead, without the need to specify the function form Φ of the non-linear mapping from the input space \mathcal{X} to the feature space \mathcal{W} .

8.4.2 Sampling via Reproducing Kernels

Kernels provide a convenient tool for sampling – they provide a means of dealing with input samples by appropriate interpolation.

8.4.2.1 Shannon Sampling

Recall that in RKHS, the sinc kernel provides a way for sampling and exact reconstruction as long as the input space and its sampling each satisfies some conditions (in terms of band-limitedness and Nyquist rate, respectively). Kernels in RKBS have similar utilities. In fact, the following Shannon sampling theorem type reconstruction formula has been established:

Theorem 8.8 (Zhang & Zhang, 2011) *Let \mathcal{B} be an RKBS on \mathcal{X} with the semi-inner product reproducing kernel G and let $\{x_j : j \in \mathbb{I}\}$ be a sequence of sampling points in \mathcal{X} . If $G(x_j, \cdot)$ and $G(x_j, \cdot)^*$ form an X_d -Riesz basis and X_d^* -Riesz basis for \mathcal{B} and \mathcal{B}^* , respectively, then the standard reconstruction operator $\mathcal{S} : \mathcal{B} \rightarrow \mathcal{B}$ defined by*

$$(\mathcal{S}f)(x) := \sum_{j \in \mathbb{I}} f(x_j) G(x_j, x)$$

is bijective and bounded. Furthermore,

$$f(x) = \sum_{j \in \mathbb{I}} f(x_j) (\mathcal{S}^{-1} G(x_j, \cdot))(x), \quad \forall f \in \mathcal{B}, x \in \mathcal{X}.$$

8.4.2.2 Sample-based Hypothesis Space with ℓ^1 -Norm

The key challenge in bounding the excess risk $\Delta(f_{\mathcal{Z}, \lambda})$ discussed in Section 8.3.1.3 arises from the fact that the solution space $\mathcal{B}^{(n)}$ used for regularized empirical risk minimization depends on training sample inputs $\{x_i\}_{i=1}^n$, and different sets of training examples will result in different subspaces $\mathcal{B}^{(n)}$ of the RKBS \mathcal{B} . This problem

has been studied (Shi, Feng, & Zhou, 2011; Wu & Zhou, 2008) and is referred to as *sample dependent hypothesis space*.

Clearly, the infinite-dimensional L^1 -spaces are not reflexive. However, we are not seeking to represent all the continuous linear functionals by kernel functions, but only continuous point evaluation functionals.

Suppose that \mathcal{X} is a metric space. To construct a reproducing kernel, we require K to be such that for each $x \in \mathcal{X}$, $K(\cdot, x)$ belongs to $C_0(\mathcal{X})$, the space of continuous functions on \mathcal{X} that vanish at infinity, and satisfies the denseness property

$$\overline{\text{span}}\{K(\cdot, x) : x \in \mathcal{X}\} = C_0(\mathcal{X}). \quad (8.20)$$

We next consider the embedding of signed Borel measures on \mathcal{X} into \mathcal{B} as induced by K . Specifically, we construct our space as

$$\mathcal{B} := \left\{ f_\mu := \int_X K(t, \cdot) d\mu(t) : \mu \text{ is a Borel measure on } \mathcal{X} \right\}$$

and define the norm of f_μ in \mathcal{B} as the total variation of the corresponding measure μ – the set of functions given by Equation (8.10) lies in \mathcal{B} and corresponds to μ taking discrete measures $\sum_{j=1}^n c_j \delta_{x_j}$ on a countable support in \mathcal{X} , whose total variation is exactly the ℓ^1 -norm of the coefficient (8.11). Therefore, \mathcal{B} indeed possesses the ℓ^1 -norm as desired. Kernel embedding of probability measure is a hot research topic (Fukumizu, Lanckriet, & Sriperumbudur, 2011; Sriperumbudur *et al.*, 2010), although in Sriperumbudur *et al.* (2011), measures are embedded into \mathcal{B}^* , the dual of \mathcal{B} (different from the approach discussed here).

This construction allows the introduction of the bilinear form on $\mathcal{B} \times C_0(\mathcal{X})$

$$(f_\mu, g) := \int_X g(t) d\mu(t), \quad f_\mu \in \mathcal{B}, \quad g \in C_0(\mathcal{X}).$$

Since the dual of $C_0(\mathcal{X})$ is precisely the set of Borel measures $M(\mathcal{X})$, K is the reproduced kernel of both \mathcal{B} and $C_0(\mathcal{X})$ in the sense that

$$f(x) = (f, K(\cdot, x)), \quad g(x) = (K(x, \cdot), g) \quad \text{for all } f \in \mathcal{B}, \quad g \in C_0(\mathcal{X}), \quad x \in \mathcal{X}.$$

Song, Zhang, and Hickernell (2013) show that \mathcal{B} satisfies the representer theorem if and only if

$$\|K[\mathbf{x}]K_{\mathbf{x}}(\cdot)\|_{\ell^1} \leq 1 \quad \text{for all pairwise distinct } \mathbf{x} = \{x_j \in \mathcal{X} : j = 1, \dots, n\}, \quad (8.21)$$

where $K[\mathbf{x}]$ is the matrix $[K(x_j, x_k) : 1 \leq j, k \leq n]$ and $K_{\mathbf{x}}(\cdot) := (K(\cdot, x_j) : 1 \leq j \leq n)^T$. It turns out that the Shepard (exponential) kernel satisfies the above relation, while the Gaussian kernel does not.

8.4.3 Kernel Method as a Unifying Feature and Similarity

Kernels provide unifying perspective on similarity-based generalization and feature-based generalization. We view similarity and feature representation as two

sides of the categorization “coin,” which can be unified via an RKBS framework. Using the semi-inner product operator tool, RKBS provides a unified treatment of regularized learning to deal with similarity-based generalization, using Hilbert-space 2-norm as a regularizer, and to deal with sparsity-based feature representation and selection, using 1-norm for regularization. According to this viewpoint, a psychological stimulus will have *dual* representations: one in the input space (where processing is described by exemplars and their similarity), and the other in the feature space (where processing is described by feature–outcome associations). So the two main issues of regularized learning, kernel design and feature selection, are heavily intertwined, with the former addressed through kernel learning in multi-task settings and the latter addressed in representational learning in sparse encoding settings.

8.4.3.1 Dictionary Learning and Feature Selection

A key element in representational learning is to acquire a set of basis functions in infinite-dimensional function space. Neurophysiological evidence suggests that setting up such a proper representational basis is accomplished in brain areas for primary sensory representation, and through means of sparse encoding. For example, in Olshausen and Field (1996), a sparse representation c of an image y is generated under an internal representation $u = \{u_k \in \mathcal{W} : k \in \mathbb{I}\}$ (where \mathbb{I} is an index set) via

$$\tilde{L}(u, x_j) := \min_{c \in \mathbb{R}^n} L\left(\sum_{k \in \mathbb{I}} c_k u_k, x_j\right) + \lambda \|c\|_{\ell^1}, \quad (8.22)$$

where the ℓ^1 -norm $\|c\|_{\ell^1} := \sum_{k \in \mathbb{I}} |c_k|$ is employed as the regularizer. The internal representation, or “dictionary” u , on the other hand, is achieved through averaging over an ensemble of input images $\mathbf{x} = \{x^j \in \mathcal{X} : j = 1, \dots, n\}$

$$\min_u \frac{1}{n} \sum_{j=1}^n \tilde{L}(u, x_j).$$

To guarantee that the dictionary $u = \{u_k : k \in \mathbb{I}\}$ is matched to the statistics of the input stimuli, the algorithm alternates between minimizing Equation (8.22) with respect to the coefficients (i.e., c_k) given the current dictionary for an example data set, then taking a gradient descent step to update dictionary elements (i.e., u_k , $k \in \mathbb{I}$):

$$u_l \leftarrow u_l + \epsilon \left\{ c_l \left(x_j - \sum_{k \in \mathbb{I}} u_k c_k \right) \right\} \quad (8.23)$$

where ϵ is the step size and $\{\cdot\}$ indicates average over the current batch of data. Using a similar dual minimization scheme, LeCun’s “convolution network” (Kavukcuoglu *et al.*, 2010), which has been very successfully applied to object recognition in vision, applied this unsupervised method for learning a multi-stage hierarchy of sparse features.

8.4.3.2 Maximal Margin Classifier and SVM

Once the set of basis functions of a feature space is established, classification is achieved by mapping input samples to points in the feature space – those from distinct categories are separated by a hyperplane serving as the “decision boundary.” The distance from the decision boundary for each example constitutes the “margin,” and it changes as one moves the placement and orientation of the separating hyperplane around. Maximal margin classifier is a learning algorithm that seeks to produce as large a margin as possible given input data, which would lower generalization error according to statistical learning theory. An example of such a classifier is the support vector machine (SVM) which, through training, establishes a set of “support vectors” or training examples which contribute to the classifier.

SVM as an application of kernel methods in Hilbert space is well established. Der and Lee (2007) applied semi-inner products to the study of maximum margin classifiers in Banach spaces. Zhang, Xu, and Zhang (2009) investigated soft margin hyperplane classification

$$\inf_w \left\{ \frac{1}{2} \|w\|_{\mathcal{W}}^2 + \lambda \|\xi\|_{\ell^1} : w \in \mathcal{W}, \xi := (\xi_j \geq 0 : 1 \leq j \leq n), b \in \mathbb{R} \right\} \quad (8.24)$$

subject to

$$y_j([\Phi(x_j), w]_{\mathcal{W}} + b) \geq 1 - \xi_j, \quad j = 1, 2, \dots, n.$$

Here, Φ is a chosen feature map from \mathcal{X} to the feature space \mathcal{W} , which is an RKBS with an semi-inner product $[\cdot, \cdot]_{\mathcal{W}}$, and λ is a fixed positive constant controlling the trade-off between margin maximization and training error minimization. The minimizer w_{opt} of Equation (8.24) was shown to belong to the closed convex cone spanned by $y_j \Phi(x_j), j = 1, 2, \dots, n$, that is,

$$w_{opt} = \sum_{j=1}^n c_j y_j \Phi(x_j)$$

for some non-negative constants c_j 's.

8.4.3.3 Kernel Learning and Vector-valued Maps

To provide a comprehensive framework for kernel learning and representational learning, the reproducing kernel Hilbert space methods can be extended to deal with learning of vector-valued functions (Micchelli & Pontil, 2005a,b), as motivated by the multi-task learning paradigm where the same input is being used for multiple objectives for inference. This vector-valued RKHS framework can be combined with the scalar-valued RKHS theory to provide a flexible framework for learning internal/feature representation (see also Micchelli & Pontil, 2007). Our work (Zhang & Zhang, 2013) has extended the vector-valued framework to the RKBS setting, which allows more explicit integration of feature and input spaces.

One such possible integration is illustrated below. We define a scalar-valued function \tilde{f} on the extended input space $\tilde{\mathcal{X}} := \mathcal{X} \times \Lambda$:

$$\tilde{f}(x, u) := [f(x), u]_{\Lambda}$$

where $\tilde{f} \in \tilde{\mathcal{H}}$ constitutes a new function space (i.e., treating feature space as a portion of the input space). Then, applying the standard (scalar-valued) RKBS theory to $\tilde{\mathcal{H}}$, the scalar-valued semi-inner product kernel function $\tilde{G} : \tilde{\mathcal{X}} \times \tilde{\mathcal{X}}$ for this space $\tilde{\mathcal{H}}$ can be characterized as (see corollary 3.3 in Zhang & Zhang, 2013)

$$\tilde{G}((x, u), (x', v)) = [\tilde{\Phi}(x, u), \tilde{\Phi}(x', v)]_{\Lambda}$$

where $\tilde{\Phi}$ is a vector-valued (as opposed to operator-valued) feature map. The representer theorem adopts the form

$$f_{opt}^* \in \overline{\text{span}} \left\{ (G(\cdot, x_j)u) : j = 1, \dots, n, u \in \Lambda \right\}.$$

If we choose $\tilde{\Phi}(x, u) = \mathcal{W}(x)u$, then the optimal solution \tilde{f}_{opt} satisfies $\tilde{f}_{opt}(x, u) = [f_{opt}(x), u]_{\Lambda}$ where f_{opt} is the solution to the vector-valued RKHS problem. The kernel

$$G(x, x') := \tilde{G}((x, f_{opt}(x)), (x', f_{opt}(x')))$$

is the optimal semi-inner product kernel for learning $\tilde{f}(x, f(x)) = \|f(x)\|_{\Lambda}^2$ with a feature map f . The vector-valued RKBS framework provides a unifying language and potential tool for addressing kernel learning and representational learning (Zhang & Zhang, 2013).

8.5 Psychological Underpinnings of Kernel Methods

8.5.1 Computational Models of Human Categorization

Category learning in cognitive psychology refers to how people acquire new concepts/categories. Typically, subjects are presented with input stimuli (a set of images, words, etc.) sequentially, classifying each and then receiving corrective feedback. A central question here concerns the mental representations people construct when acquiring new categories and storing old examples, and to perform categorical decisions. Various mathematical/computational models have been investigated by cognitive psychologists, which fall into the following broad classes.

- (i) *Exemplar models*, which postulate that people judge the similarity of a test item to all remembered exemplars of each category. Medin and Schaffer's (1978) context model and Nosofsky's (1986) generalized context model belong to this class.
- (ii) *Prototype models*, which postulate that people judge the similarity of a test item to a prototype of each category (Reed, 1972), where the prototype represents the average feature values of all category members (e.g., Smith & Minda, 1998).

- (iii) *Decision-bound models*, which postulate that people learn boundaries in stimulus space that separate categories. These were proposed by Ashby and his colleagues (Ashby & Gott, 1988; Ashby & Perrin, 1988; Maddox & Ashby, 1993).
- (iv) *Connectionist models*, which postulate that people learn associations between individual stimulus features and category labels with an error-driven mechanism (Gluck & Bower, 1988). In the ALCOVE model (Kruschke, 1992), associative learning is coupled with learning to selectively attend to relevant dimensions as individual exemplars are presented.
- (v) *Bayesian models*, which postulate that people perform Bayesian inference based on combining prior assumptions about distributions of categories and features with feature representation of an object. Categories are represented as mixture distributions, with each component (“cluster”) of this mixture given by a Gaussian distribution centered on some hypothetical stimulus. These so-called rational models of categorization were first proposed by Anderson (1991), followed up in Love, Medin, and Gureckis (2004), and popularized recently with elaborative statistical techniques by Griffiths and his associates (Austerweil & Griffiths, 2011, 2013; Sanborn, Griffiths, & Navarro, 2010).

We review each of these classes of categorization models below.

8.5.1.1 Exemplar Models

Exemplar models of categorization postulate that when a test stimulus is encountered, its similarities to the memory representation of every previously seen exemplar from each potentially relevant category are computed, and then the test stimulus is assigned to the category for which the sum of these similarities is greatest. The starting point for exemplar models is item-to-item similarity η_{ij} forming entries of a confusability matrix. The η 's are non-negative numbers, typically obtained experimentally, with η_{ij} reflecting a human subject's judgment about similarity between item i and item j . With the assumption of $\eta_{ij} = \eta_{ji}$ and $\eta_{ii} = 1$, the similarity measure is made to be monotonically related to psychological distance through the multidimensional scaling technique (Shepard, 1957), and used in Luce's (1959) choice model.

Medin and Schaffer (1978), in their highly influential paper, started out by recognizing the following important characteristics of human categorization. Natural categories do not have well-defined rules or fixed boundaries separating them, and many natural concepts cannot be defined by a single set of critical features (Rosch, 1973). Rather, members vary in the degree to which they are judged to be good (“typical”) examples of the category, and items judged to be typical of a category possess features that are characteristic of the class but not necessary for defining the category – an example being that the robin is a typical member of the category bird and has the characteristic feature that it flies, but not all birds fly, e.g., penguins. Based on these observations, Medin and Schaffer (1978) proposed (and tested) the so-called context theory of categorization,

in which similarity information between exemplars were used for classification judgments.

Nosofsky (1986), in his generalized context model, formally linked the context model with the Luce–Shepard conceptualization of identification/generalization to propose that the same stimulus similarity structure, revealed through the multi-dimensional feature representation, is used for stimulus identification and stimulus classification. Psychologically, this amounts to assuming that individuals store in their memories category exemplars which, during the categorization process, then become retrieved (not necessarily consciously) for comparison with the test stimulus to be classified. Formally, the probability of classifying a stimulus i in category J is written as

$$\text{Prob}(J|i) = \frac{b_J \sum_{j \in J} \eta_{ij}}{\sum_{K=1}^m (b_K \sum_{k \in K} \eta_{ik})}.$$

Here, upper-case symbols J, K refer to categories, and lower-case symbols i, j, k refer to individual exemplars in the corresponding category. The parameters b_J, b_K represent the biases in making a categorization response in favor of J, K , etc.

With respect to memory retrieval process, Nosofsky and Palmeri (1997) further proposed a random walk mechanism to describe speeded categorization: exemplars race among one another to be retrieved from memory, with rates determined by their similarity to a test item. The retrieved exemplars provide incremental information that enter into a random walk process for making classification decisions. In this way, the categorization process is linked with memory retrieval process explicitly. Earlier, Hintzman (1986) offered an exemplar model of episodic memory retrieval – the model, called MINERVA 2, was able to retrieve an abstracted prototype of the category when cued with the category name, and to retrieve category name when cued with an exemplar.

8.5.1.2 Prototype Models

Categorizing objects into psychological equivalence classes is historically thought of as prototype-based – humans supposedly average their experience with various exemplars to form a “prototype,” or most typical member, of a category, compare new items to it, and render category membership judgment based on how similar the item is to this prototype. One common way of defining the prototype is as the centroid of all instances of the category in some psychological space.

Reed (1972) conducted experiments to determine how people make classifications when categories are defined by sets of exemplars and not by logical rules. Using schematic faces as stimuli and college students as subjects, it was concluded that the predominant strategy was to abstract a prototype representing each category and to compare the distance of novel patterns to each prototype, emphasizing those features which best discriminated the two categories.

In response to the emerging exemplar-based models of categorization (see Section 8.5.1.1), which were directly pitted against Reed’s prototype model, Smith and Minda (1998) analyzed human performance when participants learned both

smaller, less differentiated categories and larger, more differentiated categories. They found that while the performance on former categories was better accounted for by an exemplar model, performance on the latter was better accounted for by a prototype model. Subjects' categorization strategies also differed at successive stages in learning these two types of categories: in the former case, the exemplar model dominated even early in learning, whereas in the latter case, the prototype model had a strong early advantage that gave way slowly. There seems to be a psychological transition, from *prototype-based* to *exemplar-based* processing during category learning. Smith and Minda (2000) further drew into question the particular experimental paradigm and category structure used to produce results that supported exemplar models.

8.5.1.3 Decision-bound Models

The decision-bound theory of categorization (mainly by Ashby and his associates) assumes that subjects partition the stimulus space into response regions in a multidimensional perceptual space, with these regions separated by boundaries called "decision bounds." When presented with a test stimulus, the subject determines which region the percept is in, and then emits the associated response. This class of models is closely related to the *signal detection theory* (Green & Swets, 1966), the remarkably successful theory of human perceptual detection and discrimination.

Ashby and Gott (1988), as an application of *general recognition theory* (Ashby & Townsend, 1986), recognized that subjects often used deterministic decision rules despite considerable variations in exemplars across different attributes. Subjects' use of decision rules could be rather complex (e.g., quadratic rather than linear decision boundaries), and performed nearly optimally as prescribed by the *signal detection theory*. Ashby and Perrin (1988) demonstrated that a perceptual stimulus can indeed be represented as a point in a multidimensional space, and that similarity among two stimuli is a function of the overlap of their perceptual distributions. A stochastic generalization of the static point representation of the percept to that with a multivariate diffusion process was put forward in Ashby (2000), where the decision process was modeled as a variable (distance-to-bound) driving a univariate diffusion process with two absorbing barriers.

To compare exemplar-based models with the decision-bound model, Maddox and Ashby (1993) applied sophisticated model selection and model comparison measures, and reached the following conclusions: (i) when the exemplars from each category were normally distributed and the optimal decision bound was linear, the deterministic exemplar model and the decision-bound model provided roughly equivalent accounts of the data; (ii) when the optimal decision-bound was non-linear, the decision-bound model provided a more accurate account of the data than did exemplar models. When applied to categorization data collected by Nosofsky (1986) in which the category exemplars were not normally distributed, the decision-bound model provided excellent accounts of the data, in many cases significantly outperforming the exemplar models.

8.5.1.4 Connectionist Models

Neural networks provide a general modeling framework for achieving arbitrary input–output mapping, so naturally this framework has been used to model human categorization performance. Gluck and Bower (1988) built what is essentially the Perceptron architecture (Rosenblatt, 1957) to model categorization data. The adopted architecture had one layer of distinct input units (one per physical stimulus) feeding activation directly into one of the two output units corresponding to the output categories. The training phase involved setting the connection weights using the Rescorla–Wagner (1972) rule minimizing the least mean squares (LMS) error. Because of the non-linear transfer function, this simple, two-layered network performed nonlinear discriminant analysis on input stimuli: optimal classification was achieved by predicting a criterion variable for the two categories from a set of independent variables. The network did not involve any intermediate hidden layer, so it did not need to invoke the celebrated backpropagation algorithm for error-driven learning of connection weights to hidden nodes. To conclude, Gluck and Bower (1988) showed how a computationally simple architecture (namely, Perceptron) with an error-driven learning rule accounted for empirical data of human categorization.

A major enhancement to the connectionist (Perceptron) model of category learning is Kruschke's (1992) ALCOVE (Attention Learning Covering Map) model, which incorporated both exemplar-based representation in exemplar models and error-driven learning of Perceptron. ALCOVE is a feed-forward connectionist network with three layers of nodes. Each input-layer node encodes a single psychological dimension; each hidden-layer node represents one exemplar, i.e., a particular conjunction of input dimensions in multidimensional stimulus space. It was assumed that the input spread by the hidden nodes in the model “cover” the entire multidimensional psychological space.

This three-layered connectionist network has two kinds of connection weights. The connection between the hidden layer and the output layer is essentially the two-layer perceptron architecture: the output nodes are simply weighted averages of hidden-layer activations, implementing the core idea of the exemplar model:

$$a_k^{out} = \sum_j w_{kj} a_j^{hid}.$$

Response selection is via the Luce–Shepard choice rule using a_k as choice propensity.

The connection weights from input nodes (features) to hidden nodes (feature conjunctions, which are viewed as exemplars), on the other hand, are modeled by Shepard's universal generalization kernel function, and these connections are gated by attentional weights (as separate input). So activation of hidden nodes can be written as

$$a_j^{hid} = \exp[-c(\sum_i \alpha_i |h_{ji} - a_i^{in}|^r)^{q/r}],$$

where a_j^{hid}, a_i^{in} are activation values of input node i (test stimulus) and of hidden node j (each j representing an exemplar), with α_i attention weights on input node

(dimension), $r = 1$ for city-block (ℓ^1) metric and $q = 1$ for exponential similarity. The hidden nodes therefore can be thought of as localized “receptive fields” in the multidimensional feature space. The dimensional weights α_i , which act multiplicatively on corresponding dimensions, serve to independently stretch and shrink each dimension of the input space so that across-categories stimuli are better separated while within-category stimuli are better concentrated.

Compared with standard feed-forward connectionist networks implementing parallel distributed processing (PDP), the key mechanism introduced by ALCOVE is attentional modulation of hidden nodes, which allows selective enhancement of relevant stimulus dimensions and sensitive tuning to correlated dimensions. Learning occurs as the change of connection strengths from the hidden-to-output layer, via error-driven learning, as in Gluck and Bower’s implementing of Perceptron architecture. Error backpropagation is used to modify the attention weights α_i to exemplars, in addition to modifying the weights w_{ij} . Therefore, adjustment of attention weights is specific to each input dimension separately. In particular, it will allow the model to attend to relevant features and ignore irrelevant features. Looking at the time-course of the hidden layer representation will determine which dimensions of the given representation are most relevant to the task and how strongly to associate exemplars with categories.

ALCOVE has some nice computational properties. The use of exemplars as hidden nodes allows them to interact during learning, because error spreads in a non-linear fashion to connection strengths from individual feature dimensions. The interactive nature in the learning is comparable to the Perceptron model, which accounts for the base-rate neglect phenomenon and certain forgetting effects. This interaction also implies that similar exemplars from the same category should enhance each other’s learning – this further implies that prototypical exemplars should be learned faster than non-typical exemplars that lie at the category boundary, and that the shape of category boundary, compared with clustering of exemplars, would have little direct influence on the difficulty of distinguishing categories.

8.5.1.5 Bayesian Models

Bayesian models of categorization address a slightly different problem, that of category learning or induction, that is, how category structures are created in the first place. While other computational models all assume *a priori* a fixed number of categories, Bayesian models focus on how clustering of objects occurs from probabilistic representation of their features based on some prior distributions reflecting statistical regularities of the environment.

Anderson (1990, 1991) presented a rational model of human categorization which assumed that categorization reflected the derivation of optimal estimates of the probabilities of unseen features of objects. He performed a Bayesian analysis of what optimal estimations would be if categories formed a disjoint partitioning of the space and if features were independently displayed within a category, and proposed an incremental algorithm for calculating probabilistic classification. The prior probability a new object is classified into category k :

$$\text{Prob}(k) = \frac{cn_k}{(1 - c) + c \sum_k n_k} \quad (8.25)$$

where n_k is the number of existing objects in category k , and c is a coupling constant which reflects a fixed probability that any two objects come from the same category. The prior probability that an object comes from an entirely new category not seen thus far is given by:

$$\text{Prob(new)} = \frac{1 - c}{(1 - c) + c \sum_k n_k}. \quad (8.26)$$

The likelihood function in Anderson's model, on the other hand, is expressed as the product of conditional probabilities through a Dirichlet process, assuming independence of features. For any feature dimension, let $c_i, i = 1, \dots, J$ be the number of objects of a particular category showing i -indexed value (as one of the possible J values) on that feature dimension, and α_i is the prior probability of this value distribution for that category. The conditional probability that any object of that category displays the i -indexed value is

$$\frac{\alpha_i + c_i}{\sum_j (\alpha_i + c_i)}.$$

Such a likelihood function is then combined with prior probability of class membership (8.25) and (8.26) to generate *a posteriori* classification of a new object. Anderson's (1991) rational model was shown to demonstrate many desired properties, such as effects of central tendency of categories, extraction of basic-level categories, base-rate effects, probability matching and trial-by-trial learning. Anderson's model is an example of the *Dirichlet process mixture model* in non-parametric Bayesian statistics.

Sanborn, Griffiths, and Navarro (2010), following Anderson (1991), further investigated efficient computation implementation of this rational scheme of probabilistic inference, using Markov chain Monte Carlo (MCMC) algorithms which approximate a probability distribution with a set of samples from that distribution. The advantage is twofold: the MCMC algorithms they investigated (both the Gibbs sampler version and the particle filters version) asymptotically approximate ideal Bayesian inference, with the quality of the approximation guaranteed. Second, they teased apart the underlying statistical model from the inference algorithm, therefore allowing these algorithms to be viewed as process models instead of just rational (normative) models. With respect to the empirical finding that people seem to shift from using a prototype representation early in training to using an exemplar representation late in training (Smith & Minda, 1998), Sanborn, Griffiths, and Navarro's (2010) model is able to explain such a shift as a rational statistical inference. Overall, Bayesian models cast category learning as a problem of density estimation: determining the probability distributions associated with different category labels from sequentially obtained data streams while assuming some kind of feature representation of individual objects.

To address how feature representation arose to begin with, Austerweil and Griffiths (2011, 2013) presented a computational framework for flexibly constructed

feature representations for a set of observed objects. Their non-parametric Bayesian model learned a feature representation from the raw sensory data of a set of objects without specifying the number of features ahead of time. There was potentially an infinite number of features to represent the set of objects, but representations with higher number of features were penalized. Formally, the problem of feature learning reduces to a matrix factorization problem, in which the feature-ownership matrix and feature-input matrix were simultaneously being sought.

Treating exemplar and prototype as two extremes, a number of models have explored possibilities of using clusters (each consisting of exemplars) to represent categories. Vanpaemel and Storms (2008) in the varying abstraction model (VAM) formalized a set of interpolating models by partitioning instances of each category into clusters, where the number of clusters per category could range from 1 to the total number of exemplars belonging to that category. Love, Medin, and Gureckis (2004) in the SUSTAIN (Supervised and Unsupervised Stratified Adaptive Incremental Network) model treated the discovery of category substructure as affected not only by the structure of the world but also by the nature of the learning task and the learner's goals. Its neural network model recruited new nodes for its intermediate representations of categories whenever the current architecture failed to adequately capture the input–output pattern.

8.5.2 Unified Account of Psychological Categorization Models

The above-mentioned psychological models of categorization, while emphasizing different aspects of empirical phenomena of human categorization and category learning, reveal only partial pictures of the categorization “elephant.” In this section we will review the interconnection of those psychological models of categorization, and investigate how the core ideas behind those models can be stated more precisely using terminologies from reproducing kernel methods and sparsity methods.

8.5.2.1 Relations Among Various Categorization Models

Ashby and Maddox (1993) examined several categorization models in terms of representations of stimulus and of category as well as processes for memory retrieval and for response selection. They showed some relationships between the probabilistically weighted prototype model and the decision-bound model, and how decision boundaries can also arise in exemplar-based models in general. That analysis, however, was restricted to categories with normally distributed attribute values.

Ashby and Alfonso-Reese (1995) showed equivalence of exemplar and of prototype models of categorization to different forms of density estimation. Identifying category similarity η_{iJ} with the probability of generating an item i from category J , and category bias B_J as the prior probability of category J , led to the interpretation of exemplar models as kernel density estimation and prototype models

as parametric density estimation (of the prototype parameters), both as rational solutions to the problem of categorization.

8.5.2.2 Unified Account by Reproducing Kernels

Casting psychological models of categorization (exemplar, prototype, decision-bound, connectionist, and to some extent Bayesian models) into the unifying language of kernel methods and regularized learning will further provide a coherent picture of these empirically driven modeling efforts.

The exemplar model of categorization relies on a similarity function, which is naturally conceptualized as the kernel function in machine learning (Jäkel, Schölkopf, & Wichmann, 2008a). Below, we fully elaborate the implications of such identification to human categorization models.

A test item, x , is classified according to a weighted sum of similarities to the training examples x_i : $f(x) = \sum_i c_i K(x_i, x)$. The exemplar weights, c_i , can be set to equal to the corresponding category labels, i.e. $c_i = y_i \in \{-1, 1\}$, as in Nosofsky (1986), or they can be learned to minimize classification error, as in Kruschke (1992). Reproducing kernel theory provides a key relation (8.16) between kernel similarity and feature representation under RKHS:

$$\sum_i c_i K(x_i, x) = \left\langle \sum_i c_i \Phi(x_i), \Phi(x) \right\rangle. \quad (8.27)$$

Interpreting the term $\sum_i c_i \Phi(x_i) \equiv \alpha$ as a hypothetical stimulus (“prototype”) in feature space, and $\langle \alpha, \Phi(x) \rangle$ as projections of the feature representation of stimulus x onto this prototype feature vector, then the exemplar model, which uses similarity in the input space (\mathcal{X}), becomes equivalent to a weighted prototype model represented by the prototype feature vector α in the feature space. Alternatively, we can view α as a vector of association weights, where each vector component represents the weighting of feature dimensions, and turn the exemplar model into a Perceptron model, or to a linear decision-bound model. In particular, the maximal margin classifier, in which class boundary is sought after to separate the two classes of input stimuli as far apart as possible, bears resemblance to the decision-bound model of categorization.

For non-parametric Bayesian models of categorization, we can view this approach as similar to seeking sparsity in the vector c of exemplar weights, and hence bearing some resemblance to sparse SVM with hinge loss (where sparsity of support vectors results). However, the support vectors usually lie at the boundaries of the separating hyperplanes – they are hard-to-classify exemplars or “outliers.” This is in sharp contrast to cluster models which seek to identify a small set of exemplars sufficient for representing the category structure.

Note that in the RKBS framework, the relationship generalizing Equation (8.27)

$$\sum_i c_i G(x_i, x) = \left[\sum_i c_i \Phi(x_i), \Phi(x) \right] = [\alpha, \Phi(x)]$$

still holds, due to linearity of semi-inner product with respect to the first argument. We can thus define $\alpha = \sum_i c_i \Phi(x_i)$, where exemplar weights c_i are task-dependent and feature representation $\Phi(x_i)$ is task-independent. The α -vector defined above represents the feature dimension weights for the prototype. The interaction between exemplar weight c -vector and feature weight α -vector will be of fundamental importance. Our analysis shows that the equivalence between the exemplar model and the prototype model is valid not only when perceptual “dimensions” are integral (Hilbert space), but also where feature representation is explicitly constructed (for ℓ^1 -space, a Banach space).

8.5.2.3 Shepard Kernel and Multidimensional Input Scaling

Similarity plays a central role in psychological models of category learning, as it interacts with attention and generalization. In his seminal work on generalization, Shepard (1957, 1987) argued that generalization between stimuli is a direct function of their perceived similarity. There is a long tradition in multidimensional scaling (MDS) of modeling similarity as a decreasing function of distance in some metric stimulus space. For instance, the dominant assumption is that pairwise similarity $K(x, x')$ between two stimuli x, x' is given by

$$K(x, x') = e^{-d_p(x, x')^p} \quad (8.28)$$

where d_p is the Minkowski (ℓ^p) metric over the input vector space \mathcal{X}

$$d_p(x, x') = \left(\sum_j |x_j - x'_j|^p \right)^{1/p} \quad (8.29)$$

with $p \geq 1$ and j ranges over the dimensions of the stimulus space, \mathcal{X} .

Shepard (1987) offered a normative justification for $p = 1$ (exponential kernel) based on Bayesian inference, under a model in which meaningful outcomes are associated with contiguous *consequential regions* of the stimulus space. Others have argued for $p = 2$ (Gaussian kernel), based on human performance in categorization tasks (Nosofsky, 1986). This view of psychological similarity maps naturally onto kernel methods, and indeed several psychological models founded on similarity, including models of categorization, can be formally recast as kernel methods (Jäkel, Schölkopf, & Wichmann, 2008a,b, 2009).

The standard non-metric MDS framework models stimulus similarity as a non-increasing function of the symmetric proximity data $\{K(x_i, x_j) \equiv K_{ij}, i, j = 1, \dots, n\}$ on pairwise comparison in stimulus space \mathcal{X} . In traditional studies, proximity data (K_{ij} as a matrix) are always assumed to be symmetric or are symmetrized before being fed as inputs to MDS. This is because MDS seeks to embed the set of n input stimuli *metrically* into some feature space such that proximity is monotonically related to pairwise distance. Of course, when embedded into a Hilbert space, proximity is also monotone-related to the inner product (and hence the kernel function). MDS has been tackled in contemporary machine learning as a kind of manifold learning, such as local linear embedding (LLE) (Roweis

& Saul, 2000), Isomap (Tenenbaum, de Silva, & Langford, 2000), and Laplacian eigenmaps (Belkin & Niyogi, 2003).

With respect to the Shepard kernel (8.28) with Minkowski metric (8.29), a consistent empirical finding is that separable psychological dimensions are best fit by $p = 1$ or sometimes $p < 1$, whereas integral psychological dimensions are fit by p closer to 2 (Garner, 1974; Shepard, 1964). Various proposals have been made for how learning transforms a stimulus space from having an integral representation ($p = 2$, Gaussian kernel) to a separable representation ($p = 1$, exponential kernel), where categories tend to be aligned with a consistent set of axes in the stimulus domain.

8.5.2.4 Incorporating Attention into Kernel Methods

Attention modulation of psychological similarity is a well-established principle in cognition. The pattern of pairwise similarities perceived among a set of stimuli depends on what aspects of the stimuli an observer considers currently relevant, which in turn depends on both the task and the set of stimuli present (Goldstone, 1994; Medin, Goldstone, & Gentner, 1993; Shepard, 1964; Tversky, 1977). Generalization is stronger between stimuli differing on task-irrelevant dimensions than between stimuli differing on task-relevant dimensions (Jones, Maddox, & Love, 2005; Nosofsky, 1986).

Attention might affect categorization by (i) influencing the sampling of input stimulus, through imposing weighting on individual exemplars on \mathcal{X} (input filtering), or (ii) highlighting feature dimensions in the feature space, through modifying the shape of the kernel. Mathematically, (i) corresponds to controlling the exemplar weights c_i and hence imposing a measure on \mathcal{X} , while (ii) corresponds to parameterizing the family of kernels or equivalently the family of semi-inner products of a common basis or frame (e.g., the power exponent p in the semi-inner product kernel which, in the $p = 1$ case, is simply the projection operation).

Attentional effects on perceived similarity have been modeled by assuming learnable weights for each dimension (Nosofsky, 1986). In the language of kernels, it is given by the ℓ^p metric (8.28) where $p = 1$ or 2:

$$K(x, x'; \alpha) = \begin{cases} \exp\left(-\sum_j \alpha_j |x_j - x'_j|\right) & \text{Shepard (exponential) kernel} \\ \exp\left(-\sum_j \alpha_j (x_j - x'_j)^2\right) & \text{Gaussian kernel} \end{cases} \quad (8.30)$$

where j indexes the dimension of the input space \mathcal{X} , and α_j 's are adjustable attention weights. By committing to the dimensions of the stimulus space and only leaving open their relative scaling, attention to the feature dimension serves to limit estimation error in kernel learning. In the popular ALCOVE model (Kruschke, 1992), attention weights are learned by gradient descent on a loss function.

Another conceptualizing of attention is that it controls the saliency of stimulus features, thereby influencing their learning rates (Kruschke, 2001; Mackintosh, 1975; Rescorla & Wagner, 1972). For instance, modifying the Perceptron learning

rule (Rosenblatt, 1958), theories of cue associability have proposed that the change in association weights $w = [w_1, \dots, w_m]^T$ follows gradient descent on loss $L_{\mathcal{Z}}$ over training data \mathcal{Z}

$$w_j \leftarrow w_j + \epsilon \alpha_j \frac{\partial}{\partial w_j} L_{\mathcal{Z}}(w)$$

where j is the feature dimension and α_j is the attention allocated to it. Associations from attended features are thus learned faster than associations from less-attended features. This role of attention can be shown to be equivalent to modifying the kernel by reweighting the dimensions of feature space (Matt Jones, personal communication).

Recently, stochastic techniques have been introduced to analyze ℓ^1 -regularized loss minimization (Shalev-Shwartz & Tewari, 2011). The stochastic coordinate descent method updates the weight of a single feature at each iteration; this corresponds to selective attention to feature dimensions. The stochastic gradient descent (and its generalization, stochastic mirror descent, see Srebro, Sridharan, & Tewari, 2011) method, on the other hand, involves picking randomly one example from a (uniformly distributed) training set at each iteration, and updating the exemplar weight vector based on the chosen example; this corresponds to selective attention to exemplars. These techniques are expected to generate insights for addressing the issue of attentional control in category induction (attention to feature versus attention to exemplar).

8.5.3 Challenges for Kernel Methods

Despite the conceptual advantages provided by the reproducing kernel method, there are still challenges in finding an efficient, effective, and well-understood algorithm for categorization (in contrast to state-of-the-art reinforcement learning algorithms). The key ingredient lacking is a theory for deriving effective feature representation – there are certain representational mechanisms used for human categorization that have yet to be captured by any machine learning algorithm that approaches human inductive capacity.¹

8.5.3.1 Similarity vs. Dissimilarity

Psychological similarity was known to be incompatible with the metric axioms, in that it might be non-symmetric and might violate triangle inequality (Tversky, 1977; Tversky & Gati, 1982). The general pattern observed regarding non-symmetry has been that x is rated as more similar to y than vice versa if y is more salient or has more features than does x . MDS models can account for at least some of these findings by incorporating stimulus- or response-bias parameters (Nosofsky, 1991), but non-symmetric similarity may also reflect stimulus representations themselves, as well as the process of comparison (Polk *et al.*, 2002). The RKBS method avoids the symmetry requirement for its kernels, so in principle

¹ This chapter has not touched upon deep learning, which has come into fashion since the authors worked on its first draft.

non-symmetry can be accommodated more readily, via the semi-inner product operation. However, the metric of a Banach space is still symmetric, and satisfies the triangular inequality. So it appears that semi-inner product and metric/norm capture different aspects of the geometry, although from the former the latter is derivable. In machine learning, one may simultaneously consider both “similarity” and “dissimilarity” (as in von Luxburg, 2004, for instance); in the present context, they map to, respectively, norm-induced metric structure and semi-inner product structure.

Psychological similarity is notoriously labile, depending on both the stimulus set and the task (Tversky, 1977). Much of the flexibility of human cognition arguably lies in the fact that the features underlying similarity vary in an adaptive way from task to task (Medin, Goldstone, & Gentner, 1993). This view has led to claims that it is the choice of features, and not similarity itself, that determines behavior (Goodman, 1972). Nevertheless, similarity may play an important role in cognitive processing, by providing a compact (unidimensional) indication of how well two stimuli match across a large number of features (Goldstone, 1994). From this point of view, similarity can be viewed as the mind’s implementation of the kernel trick, to perform learning and inference directly on the input space (i.e., perceptual representation), bypassing the often high-dimensional feature space. Flexibility of similarity then becomes flexibility in the choice of the kernel, K , or equivalently, of the feature map, Φ .

8.5.3.2 Flexible Feature Representation

There is also evidence that the dimensions of the stimulus space themselves can change, perhaps over longer time-scales.

To begin with, there is a long-recognized distinction between perceptually integral dimensions (2-norm) versus perceptually separable dimensions (1-norm) in the multidimensional psychological space where stimuli are represented (Garner, 1974). Moreover, perceptually integral dimensions ($p = 2$) can become separable ($p = 1$), either during the course of development or (perhaps temporarily) as a consequence of classification learning in adulthood. Goldstone and Steyvers (2001) investigated the process of dimension differentiation, whereby subjects learned a separable representation of previously integral dimensions. This change of representation entailed a qualitative shift of similarity, which affected the learnability of future categorization tasks. Jones and Goldstone (2013) showed that learning a categorization task with integral-dimension stimuli could induce an analytic coordinate frame (i.e., a set of orthogonal axes) for the stimulus space such that, when subjects were transferred to a new task, transfer performance was better if the new category boundary was aligned with this coordinate frame. When stimuli vary along integral dimensions and their distribution (within some task domain) had a non-spherical covariance structure (i.e., weighted l^2 -norm), the dimension of greatest variation (i.e., the first principle component) was shown to become perceptually highlighted. Austerweil and Griffiths (2013) reported that new stimulus dimensions could emerge after the discovery of components useful

in predicting some outcome variable. These examples illustrate how the perceptual system (mediating similarity judgments) adapts to suit the needs of the cognitive system. How such mechanisms of adaptation lead to modifying kernels to improve learning has yet to be worked out.

8.6 Summary: Unifying Feature and Similarity by RKBS

Our chapter reviews the basic concepts of reproducing kernel Banach space (RKBS) framework as potentially unifying similarity and feature representation, the two sides of the categorization “coin,” both from the mathematical perspective and from the cognitive perspective. Mathematically, RKBS will provide a unified treatment of regularized learning under both the smoothness setting (L^2 or Hilbert space) and the sparsity setting (ℓ^1 space), through the semi-inner product tool. The connection and interaction between similarity-based generalization (under 2-norm) and the sparsity-based feature selection (under 1-norm) can now be addressed under a common framework. Representer theorem continues to work so long as the regularizer is norm-based and the loss function is proper (i.e., it is non-negative, with its minimum zero achieved when $y_i = f(x_i), \forall i$). More importantly, the kernel function still depends on the norm used in the regularizer and not on the sample points. Because the Banach space approach includes the 2-norm (Hilbert space) and 1-norm as special cases, this approach has potential in bringing fundamental breakthroughs to learning, sampling, approximation, and feature representation after synthesizing kernel-based method with sparsity-based methods. Translating extant psychological models of categorization, e.g., the exemplar model, the prototype model, the decision-bound model, the connectionist model, and the cluster model, in this universal framework of kernel methods allows a more comprehensive view of psychological processes underlying categorization and category learning. A psychological stimulus may have dual representations, one in the input space (where processing is described by exemplars and their similarities) and the other in the feature space (where processing is described by feature–outcome associations). Shepard’s universal law of generalization plays the crucial role of a “kernel.” The RKBS framework using the semi-inner product will bring together similarity-based cognitive mechanisms with feature-/attention-based cognitive mechanisms for categorization, and provide a unified account of objects with dual representations, respectively, in input space and in feature space. It may well be the case that the above-mentioned psychological models merely reflect different views about whether the classification algorithm (and attention modulation) is operating on the input space or the feature space, rather than contain computationally meaningful differences in the nature of category representations (as psychologists have debated over decades)! The real questions, then, become (i) the nature of the stimulus or feature representation, and (ii) the learning algorithms that find or approximate the optimal classifier (e.g., by updating exemplar and/or feature weights). A regularized learning framework thus provides the potential for

a unification of all of these theoretical considerations, facilitating comparison and enabling new, more principled psychological models that synthesize the insights offered by each.

Acknowledgement

The authors acknowledge research support from ARO Grant No. W911NF121-0163 and AFOSR Grant No. FA9550-13-1-0025 to Jun Zhang and NSFC Grant No. 11101438 and No. 11222103 to Haizhang Zhang. We thank Matt Jones for inputs on human categorization literature and for critical reading of an earlier draft.

References

- Anderson, J. R. (1990). *Cognitive psychology and its implications*. New York, NY: WH Freeman/Times Books/Henry Holt & Co.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409–429.
- Argyriou, A., Micchelli, C. A., & Pontil, M. (2005). Learning convex combinations of continuously parameterized basic kernels. In P. Auer & R. Meir (eds.), *Learning theory. COLT 2005. Lecture Notes in Artificial Intelligence*, Vol. 3559. Berlin: Springer-Verlag.
- Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68, 337–404.
- Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, 44, 310–329.
- Ashby, F. G. & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, 39, 216–233.
- Ashby, F. G. & Gott, R. E. (1988). Decision rules in the perception and categorization of multi-dimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 33–53.
- Ashby, F. G. & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372–400.
- Ashby, F. G. & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, 95, 124–150.
- Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154–179.
- Austerweil, J. L., & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cognitive Psychology*, 63, 173–209.
- Austerweil, J. L., & Griffiths, T. L. (2013). A nonparametric Bayesian framework for constructing flexible feature representations. *Psychological Review*, 120, 817–851.
- Belkin, M. & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15, 1373–1396.
- Candès, E. J. & Tao, T. (2004). Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Transactions on Information Theory*, 52, 5406–5425.

- Candès, E. J., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52, 489–509.
- Casazza, P., Christensen, O., & Stoeva, D. T. (2005). Frame expansions in separable Banach spaces. *Journal of Mathematical Analysis and Applications*, 307, 710–723.
- Coifman, R. R. & Maggioni, M. (2006). Diffusion wavelets. *Applied Computational Harmonics Analysis*, 21, 53–94.
- Cucker, F. & Smale, S. (2002). On the mathematical foundations of learning. *Bulletin of the American Mathematical Society (New Series)*, 39, 1–49.
- Der, R. & Lee, D. (2007). Large-margin classification in Banach spaces. *AISTATS*, 2, 91–98.
- Evgeniou, T., Micchelli, C. A., & Pontil, M. (2005). Learning multiple tasks with kernel methods. *Journal of Machine Learning Research*, 6, 615–637.
- Foucart, S. & Lai, M.-J. (2009). Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *Applied Computational Harmonics Analysis*, 26, 395–407.
- Fukumizu, K., Lanckriet, G. R. G., & Sriperumbudur, B. (2011). Learning in Hilbert vs. Banach Spaces: A measure embedding viewpoint. *Advances in Neural Information Processing Systems*, 25, 1773–1781.
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum Associates.
- Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, 8, 404–409.
- Giles, J. R. (1967). Classes of semi-inner-product spaces. *Transactions of the American Mathematical Society*, 129, 436–446.
- Gluck, M. A. & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227–247.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52, 125–157.
- Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, 130, 116–139.
- Goodman, N. (1972). Seven structures on similarity. In N. Goodman (ed.), *Problems and projects*. New York, NY: Bobbs-Merrill.
- Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*, Vol. 1. New York, NY: Wiley.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93, 411–428.
- Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008a). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin and Review*, 15, 256–271.
- Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008b). Similarity, kernels, and the triangle inequality. *Journal of Mathematical Psychology*, 52, 297–303.
- Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2009). Does cognitive science need kernels? *Trends in Cognitive Sciences*, 13, 381–388.
- James, R. C. (1964). Characterizations of reflexivity. *Studia Mathematica*, 23, 205–216.

- Jones, M., & Goldstone, R. L. (2013). The structure of integral dimensions: Contrasting topological and Cartesian representations. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 111–132.
- Jones, M., Maddox, W. T., & Love, B. C. (2005). Stimulus generalization in category learning. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pp. 1066–1071.
- Kavukcuoglu, K., Sermanet, P., Boureau, Y-L., Gregor, K., Mathieu, M., & LeCun, Y. (2010). Learning convolutional feature hierachies for visual recognition. *Advances in Neural Information Processing Systems*, 23, 1090–1098.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812–863.
- Lanckriet, G. R. G., Cristianini, N., Bartlett, P., El Ghaoui, L., & Jordan, M. I. (2004). Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5, 27–72.
- Lavery, J. E. (2000). Shape-preserving, multiscale fitting of univariate data by cubic L_1 smoothing splines. *Computer Aided Geometric Design*, 17, 715–727.
- Li, S. (1995). On general frame decompositions. *Numerical Functional Analysis and Optimization*, 16, 1181–1191.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309–332.
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.
- Lumer, G. (1961). Semi-inner-product spaces. *Transactions of the American Mathematical Society*, 100, 29–43.
- Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 82, 276–298.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, 53, 49–70.
- Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.
- Micchelli, C. A. & Pontil, M. (2005a). Learning the kernel function via regularization. *Journal of Machine Learning Research*, 6, 1099–1125.
- Micchelli, C. A. & Pontil, M. (2005b). On learning vector-valued functions. *Neural Computation*, 17, 177204.
- Micchelli, C. A., & Pontil, M. (2007). Feature space perspectives for learning the kernel. *Machine Learning*, 66, 297–319.
- Nath, B. (1971/72). On a generalization of semi-inner product spaces. *Mathematical Journal of Okinawa University*, 15, 1–6.
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14, 849–856.
- Noles, N. S. & Gelman, S. A. (2012). Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches. *Developmental Psychology*, 48, 890–896.

- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94–140.
- Nosofsky, R. M. & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266–300.
- Olshausen, B. A. & Field, D. J. (1996). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7, 333–339.
- Poggio, T. & Smale, S. (2003). The mathematics of learning: Dealing with data. *Notices of the AMS*, 50, 537–544.
- Polk, T. A., Behensky, C., Gonzalez, R., & Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: Asymmetries induced by manipulating exposure frequency. *Cognition*, 82(3), B75–B88.
- Rao, M. M. & Ren, Z. D. (1991). *Theory of Orlicz spaces*. New York, NY: Marcel Dekker.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382–407.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.
- Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328–350.
- Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Roweis, S. & Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290, 2323–2326.
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117, 1144–1167.
- Schölkopf, B. & Smola, A. J. (2001). *Learning with kernels: Support vector machines, regularization, optimization, and beyond (adaptive computation and machine learning)*. Cambridge, MA: MIT Press.
- Shalev-Shwartz, S. & Tewari, A. (2011). Stochastic methods for ℓ_1 -regularized loss minimization. *Journal of Machine Learning Research*, 12, 1865–1892.
- Shawe-Taylor, J. & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge: Cambridge University Press.
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22, 325–345.
- Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, 54–87.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323.

- Shi, L., Feng, Y.-L., & Zhou, D.-X. (2011). Concentration estimates for learning with ℓ^1 -regularizer and data dependent hypothesis spaces. *Applied and Computational Harmonic Analysis*, 31, 286–302.
- Smale, S. & Zhou, D.-X. (2007). Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26, 153–172.
- Smale, S., Rosasco, L., Bouvrie, J., Caponnetto, A., & Poggio, T. (2010). Mathematics of neural response. *Foundations in Computational Mathematics*, 10, 67–91.
- Smith, E. E. (1995). Concepts and categorization. In E. E. Smith & O. Daniel (eds.), *Thinking*, Vol. 3 (pp. 3–33) Cambridge, MA: MIT Press.
- Smith, J. D. & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1411–1436.
- Smith, J. D. & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 3–27.
- Song, G., Zhang, H., & Hickernell, F. J. (2013). Reproducing kernel Banach spaces with the ℓ^1 norm. *Applied and Computational Harmonic Analysis*, 34, 96–116.
- Srebro, N., Sridharan, K., & Tewari, A. (2011). On the universality of online mirror descent. *Advances in Neural Information Processing Systems*, 24, 2645–2653.
- Sriperumbudur, B., Gretton, A., Fukumizu, K., Scholkopf, B., & Lanckriet, G.R.G. (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11, 1517–1561.
- Sriperumbudur, B., Fukumizu, K., & Lanckriet, G.R.G. (2011). Universality, characteristic kernels and RKHS embedding of measures. *Journal of Machine Learning Research*, 12, 2389–2410.
- Sutherland, N. S. & Mackintosh, N. J. (1971). *Mechanisms of Animal Discrimination Learning*, NY: Academic Press.
- Tenenbaum, J., de Silva, V., & Langford, J. (2000). A global geometric framework for non-linear dimensionality reduction. *Science*, 290, 2319–2323.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58, 267–288.
- Tong, H., Chen, D.-R. & Yang, F. (2010). Least square regression with ℓ^p -coefficient regularization. *Neural Computation*, 22, 3221–3235.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Tversky, A. & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, 89, 123–154.
- Vanpaemel, W. & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin and Review*, 15, 732–749.
- von Luxburg, U. (2004). Statistical learning with similarity and dissimilarity functions. Max Planck Institute for biological cybernetics, Tübingen, Germany. PhD thesis.
- Wu, Q. & Zhou, D.-X. (2008). Learning with sample dependent hypothesis spaces. *Computers & Mathematics with Applications*, 56, 2896–2907.
- Weiss, Y. (1999). Segmentation using eigenvectors: a unifying view. *Proceedings of the 7th IEEE International Conference on Computer Vision*, 975–982.
- Xu, Z., Zhang, H., Wang, Y., Chang, X., & Liang, Y. (2010). $L^{1/2}$ regularization. *SCIENCE China Information Sciences*, 53, 1159–1169.

- Zhang, H. & Zhang, J. (2010). Generalized semi-inner products with applications to regularized learning. *Journal of Mathematical Analysis and Applications*, 372, 181–196.
- Zhang, H. & Zhang, J. (2011). Frames, Riesz bases, and sampling expansions in Banach spaces via semi-inner products. *Applied and Computational Harmonic Analysis*, 31, 1–25.
- Zhang, H. & Zhang, J. (2012). Regularized learning in Banach spaces as an optimization problem: Representer theorems. *Journal of Global Optimization*, 54, 235–250.
- Zhang, H. & Zhang, J. (2013). Vector-valued reproducing kernel Banach spaces with applications to multi-task learning. *Journal of Complexity*, 29, 195–215.
- Zhang, H., Xu, Y., & Zhang, J. (2009). Reproducing kernel Banach spaces for machine learning. *Journal of Machine Learning Research*, 10, 2741–2775.

9 The Axiom of Meaningfulness in Science and Geometry

Jean-Claude Falmagne, Louis Narens,
and Christopher Doble

Preamble	375
Part I: The Meaningfulness Concept and the Theory of Measurement	376
9.1 History and Background	376
9.2 Dimensional Analysis	378
9.2.1 Example: Dimensional Analysis of a Simple Pendulum	379
9.3 The Representational Theory of Measurement	382
9.4 Meaningfulness in Geometry	388
9.5 Comparison of Geometric and Measurement-theoretic Concepts	391
9.6 Dimensional Analysis within the Representational Theory	393
Part II: Meaningfulness and Abstract Axiomatization of Scientific Laws	394
9.7 Meaningfulness as a Principle of Theory Construction	394
9.7.1 An Example of a Meaningfulness Condition	396
9.7.2 Examples of Abstract Axioms with their Representations	397
9.7.3 Associativity and the Pythagorean Theorem	399
9.7.4 Order-invariance under Transformations	400
9.7.5 Some Basic Concepts	401
9.8 Representation Theorems of the Seven Abstract Axioms	402
9.8.1 The Associativity Equation	402
9.8.2 Permutability and Quasi-permutability	402
9.8.3 The Translation Equation	404
9.8.4 The Bisymmetry Equation	405
9.8.5 The Autodistributivity Equations	405
9.8.6 The Abstract Lorentz–FitzGerald–Doppler Equations	406
9.9 Defining Ratio Scale Meaningfulness	407
9.9.1 Motivation	407
9.9.2 Defining Meaningfulness	409
9.9.3 Applications	411
9.9.4 The Falmagne and Narens Definition	412
9.10 Propagating Abstract Axioms via Meaningfulness	416

9.10.1 Propagating Solvability and other Conditions	417
9.10.2 The Self-transforming Case	417
9.10.3 The Meaningful Lorentz–FitzGerald–Doppler Systems	417
9.11 Meaningful Representations of Scientific Ratio-scaled Codes	419
9.11.1 Associativity and the Pythagorean Theorem	419
9.11.2 Meaningful Quasi-permutable Laws	420
9.11.3 Quasi-permutability and the Pythagorean Theorem	425
9.11.4 Meaningful Bisymmetric Laws	427
9.11.5 Meaningful Translatable Laws	428
9.12 Order-invariance under Transformations	428
9.12.1 Transformation Families	429
9.12.2 Transformation Classes	432
9.12.3 Transformations Acting on Codes	433
9.12.4 Meaningful Transformations	436
9.12.5 The Lorentz–FitzGerald Contraction	439
9.12.6 Beer’s Law	439
9.12.7 The Monomial Laws	440
9.12.8 A Counterexample: Van der Waals’ Equation	441
Part III: Meaningfulness and Scientific and Geometric Content	443
9.13 Principles for Scientific and Geometric Content	443
9.13.1 Intrinsicness	451
References	453

Preamble

Measurement is the assignment of numbers to qualitative or empirical objects, and meaningfulness is a theory for the proper use of the assigned numbers for drawing scientific inferences and conclusions. The theories of measurement and of meaningfulness are closely related, and each has important philosophical and practical implications. This chapter focuses on the concept of meaningfulness.

The chapter has three parts. Part I emphasizes the history of the meaningfulness concept and the foundational aspects of measurement theory, in particular the representational theory of measurement. Attention is given to the relationship between meaningfulness and Klein’s Erlanger program in geometry (Klein, 1872) in the late nineteenth century. Part II details applications of meaningfulness in the spirit of dimensional analysis from classical physics. These applications, largely inspired by the seminal work of Luce (1959, 1964, 1990), aim to infer the laws relating measurement variables in a scientific or geometric context. In particular, in Part II it is shown, through a number of exemplary cases, that when properly axiomatized, meaningfulness conditions paired with additional, relatively weak, intuitive invariance assumptions yield the possible functional forms of such laws, perhaps up to the values of some parameters. Part III deals with a characterization of the scientific content of a fragment of science through principles based on

meaningfulness. All three parts deal with foundational issues in science, with Part I being more historical, Part II more applied and practical, and Part III more philosophical. Parts I and III require basic concepts from (abstract) algebra, while Part II uses terminology and results from a first course in analysis, and functional equations (in the sense of Janos Aczél's basic text; see Aczél, 1966). Part III requires some familiarity with and elementary knowledge of set theory, in particular, the Zermelo–Fraenkel axiomatic approach to set theory.

Part I: The Meaningfulness Concept and the Theory of Measurement

9.1 History and Background

The term “meaningfulness” was introduced into the measurement literature in a highly influential *Science* article by the psychologist S. S. Stevens (1946). The article was a response to physicists and others who criticized Stevens' methods for establishing psychophysical scales involving subjective judgments of intensity. The critics argued that his methods – and more generally methods used in psychological measurement – could not produce the strong kinds of measurement scales needed to support rich mathematical modeling. In essence, their claim implied that psychology is necessarily a qualitative science that could only be based on weak forms of measurement, such as ordinal measurement or mere counting.

Stevens responded by proposing a new idea about what scientific measurement was. According to Stevens, all that was needed were:

- (1) a systematic rule for assigning numbers to objects;
- (2) a way of precisely describing all ways of assigning numbers to objects (i.e., determining the *scale type* of the assignment process); and
- (3) using the criteria that statistical results concerning the numerical assignments should be invariant under assignments described in (2).

He called the statistics in (3) “meaningful statistics.” He provided four examples of the kinds of scale types that could result from this form of measurement: a *nominal scale* that counts the number of objects, an *ordinal scale* that numerically orders the objects, a *ratio scale* that compares sizes of objects in terms of ratios, and an *interval scale* that compares differences in sizes of objects. Later, Stevens added two additional scale types, an *absolute scale* that assigns numbers to objects in a unique way, and a *log-interval scale* that assigns numbers in a way such that the logarithm of the numbers forms an interval scale.

Stevens' theory of measurement has obvious gaps. His concept of “a systematic rule” is vague. He only provides examples of a few scale types. Are there more? Many more? How are “meaningful” statistics different from ordinary statistics? If they are, does that mean in such circumstances that ordinary statistical methods are wrong? More serious concerns involve the use of his measurement theory: Stevens

confuses numerals with numbers, doesn't understand that although the empirical process he uses to generate measurements in his experiments generates what he call "numerals," his methods do not produce sufficiently rich structure on the numerals for them to have properties of a number system that he can apply his meaningful statistics to, etc.

On the positive side, Stevens realized that psychology needed a different foundational measurement theory than that employed in classical physics, and he provided one. Stevens' methods turned out to be highly influential in psychology, sociology, and political science, and more generally in the science of behavior. Even today there are a profusion of studies in which subjects rate items on an integer scale, e.g., 1 to 10, and the data are analyzed through use of an interval scale. If pressed for a justification, most would point to articles by Stevens or give an account that is in line with Stevens' theory of measurement.

Stevens is most famous for using his methods to argue that the psychophysical law relating subjective experience of intensity to its physical measurements is a power law. The International Society for Psychophysics calls this "Stevens' law." We will call it the *power law*¹ in this chapter.

Stevens obtained the power law by applying his method of measurement to data that consisted of subjects' numerical judgments of the ratios of subjective intensities. He assumed – or more likely, considered it part of his measurement method – that these judgments were on a ratio scale.

The mathematical psychologist R. Duncan Luce realized that the power law followed from the measurement assumptions that Stevens made if ideas involving the invariance of physical laws under changes of units (i.e., dimensional analysis) were applied to psychological laws.² This led Luce to reformulate physical dimensional analysis in a manner that made sense for psychological modeling. From this reformulation he obtained Equation (9.1), and from Equation (9.1) he was able to deduce the power law. Luce writes:

The following remarks arise from my reflections on four articles and commentaries on them. Historically, the first was my 1959 article "On the Possible Psychophysical Laws," which attempted to account for why two ratio-scaled variables, such as those encountered in the simplest version of cross-modal matching, should be related by power functions. In it, I postulated that if x and y are two ratio-scaled variables that are related by some law $y = f(x)$, where f is a strictly

1 The Belgian physicist Plateau (1872) derived the power law by mathematical methods relating physical invariances to data from experienced subjects who made bisection judgments involving gray paints. (This is very much in line with the meaningfulness methods used in Part II.) Plateau had only a few subjects and each gave one judgment. Nevertheless, Plateau's theory with these few subjects produced a strong argument for the power law.

2 Dimensional analysis is a set of techniques used in physics and engineering. They are particularly useful in very complicated physical settings where exact solutions by purely mathematical methods are unknown or perhaps even impossible. It is also indispensable in those situations, where, for all practical purposes, it is impossible to give precise and detailed formulations of the fundamental equations from which the solution is sought. A simple example of dimensional analysis is given in Section 9.2.

increasing function, and if the units of x are changed by a ratio transformation r , then there is a corresponding ratio change, $s(r)$ of y such that for all positive x and r ,

$$s(r)f(x) = s(r)y = f(rx). \quad (9.1)$$

As is easily demonstrated, this functional equation for f implies that f is a power function.

The key issue surrounding the article, which was first critically discussed by Rozeboom (1962a, 1962b), is: Why should one assume Equation (9.1)? I had spoken of it as a “principle of theory construction,” thinking it to be on a par with the dimensional invariance of physical laws postulated in the method of dimensional analysis (for a detailed discussion of that method, see Chapter 10 of Krantz, Luce, Suppes, and Tversky (1971), and Chapter 22 of Luce, Krantz, Suppes, & Tversky, (1990)). In the face of Rozeboom’s criticism, I retreated from that position (Luce, 1962). [...] Aczel, Roberts, and Rosenbaum (1986), Roberts and Rosenbaum (1986), and Osborne (1970) studied various generalizations of Equation (9.1), and Falmagne and Narens (1983) gave a detailed analysis of a collection of closely related principles, showing how they interrelate and what they imply about the forms of laws. (See Part II.) Despite these interesting developments, I have never felt that we have gained a full understanding of what is really involved empirically in assuming Equation (9.1), except I have come to recognize that it is not really the specialization of dimensional invariance that I had thought it was in 1959.

(Luce, 1990, p. 66.)

Luce’s “principle of theory construction” is discussed in the context of meaningfulness in Section 9.7 of Part II of this chapter. Also discussed there is a more detailed account of the literature that flowed from Luce’s 1959 article than that given in the above quotation. Luce was one of the founders of mathematical psychology. He thought very deeply about how to transform psychology as a science so that it more closely resembled the practices and methodology used in classical physics. He was particularly interested in understanding in a very abstract way the foundations of physical measurement, and why the methods of dimensional analysis were so powerful a tool in physics and engineering. This led him and his colleagues to generalize the foundations of physical measurement to a wide variety of measurement techniques and provide an axiomatic foundation for dimensional analysis. He finally was able to demonstrate that the key ideas from dimensional analysis were derivable from meaningfulness considerations from the theory of measurement. This part of Luce’s contributions to measurement and meaningfulness is briefly discussed in Section 9.6.

9.2 Dimensional Analysis

In order to understand Luce’s journey and its role in the theories of measurement and meaningfulness, a description of dimensional analysis is needed. A full description of its foundation with some of its powerful uses would take a whole book. Instead, a very simple example of its use is presented. What is needed to rigorously extend its use in science are:

- (i) a mathematical foundation and justification for its methods, and
- (ii) a way of generalizing the methods so that they apply outside of physics.

Sections 9.3 and 9.6 describe accomplishments that led to (i), and Part II presents our method for accomplishing (ii).

9.2.1 Example: Dimensional Analysis of a Simple Pendulum

The following is a well-known example of dimensional analysis. It follows the presentation by Narens (2002a). Consider a simple pendulum consisting of a ball suspended by a string. The ball has mass m , the string has length d , and g is the gravitational constant that describes acceleration of small masses towards the center of the earth. The ball is displaced so that the pendulum makes an angle of θ with the vertical and is then released. We want to find its period t . The letters m , g , d , and t denote *qualitative* entities which can be measured in various ways – that is, they can be properly assigned numerical values in various ways. It is assumed that they are measured in terms of a coherent set of units, i.e., if d is measured in terms of the unit u and t in terms of the unit v , then g is measured in terms of the unit u/v^2 . In this application θ is measured by the ratio of a measured length of arc divided by a measured radius, and is thus a real number – a “dimensionless quantity” – that does not depend on the unit that length is measured in. The concept is that through the use of “physical intuition,” there is enough information to determine t – that is, it is assumed that t is physically determined by m , g , d , and θ . Using functional notation, this is written as,

$$t = F(m, g, d, \theta), \quad (9.2)$$

where F is a qualitative function. A proper numerical representation for Equation (9.2) consists in giving proper numerical measurements, t , m , g , d , and θ , to t , m , g , d , and θ , respectively, through some coherent system of units and finding a numerical function F such that

$$t = F(m, g, d, \theta). \quad (9.3)$$

A fundamental principle of dimensional analysis called *dimensional invariance* requires any other proper representation for Equation (9.2) to have the form

$$t' = F(m', g', d', \theta), \quad (9.4)$$

where t' , m' , g' , and d' are proper numerical measurements in some other coherent system of units. Note that in Equations (9.3) and (9.4) the same numerical function F is used to represent the qualitative function F . The letter θ also appears in Equations (9.3) and (9.4). By the way it was defined, θ is a real number that does not depend on which unit is used to measure length, and thus it has the same numerical value for each coherent system of units. Because t , m , g , d , and θ are measured on

ratio scales, t' , m' , g' , and d' are related to t , m , g , and d in the following manner: there are positive real numbers α , β , and γ such that

$$t' = \alpha t, \quad d' = \beta d, \quad g' = \frac{\beta g}{\alpha^2}, \quad \text{and} \quad m' = \gamma m.$$

Thus by substituting these equations into Equation (9.4) we get

$$\alpha t = F\left(\gamma m, \frac{\beta g}{\alpha^2}, \beta d, \theta\right). \quad (9.5)$$

By dimensional invariance Equation (9.5) is true for all choices of positive reals α , β and γ . Since θ is fixed and its measurement does not vary with changes of units, we can rewrite Equation (9.5) as

$$\alpha t = H\left(\gamma m, \frac{\beta g}{\alpha^2}, \beta d\right), \quad (9.6)$$

where for all positive real numbers x , y , and z ,

$$F(x, y, z, \theta) = H(x, y, z).$$

Because the first argument of H in Equation (9.6), γm , can take any real value (by an appropriate choice of γ) while leaving the value of H , αt , unchanged, it follows that H does not depend on its first argument, and therefore Equation (9.6) can be rewritten as

$$\alpha t = K\left(\frac{\beta g}{\alpha^2}, \beta d\right). \quad (9.7)$$

By choosing units so that $\alpha = \sqrt{g}/\sqrt{d}$ and $\beta = 1/d$, it follows from Equation (9.7) that

$$t = \frac{\sqrt{d}}{\sqrt{g}} K(1, 1). \quad (9.8)$$

It is easy to see by inspection that when the measurements of t , g , and d are changed to another coherent system of units, an equivalent form of Equation (9.8) with the same real constant $K(1, 1)$ is valid. If it is assumed that the measurement of g is known in one coherent system of units – and thus by ratio scalability in all coherent systems of units – then Equation (9.8) gives a way of calculating for each length d the measurement of the period t in terms of the measurement of d and the real number $K(1, 1)$. The number $K(1, 1)$ cannot be determined by dimensional analysis alone. It can, however, be found by experiment; that is, a particular measured length can be chosen and the period measured for that particular length, and $K(1, 1)$ can be computed by Equation (9.8). In the above argument θ was kept fixed. As θ varies, so will Equation (9.8), but Equation (9.8) will keep the same form; that is, only the real number $K(1, 1)$ will vary. In other words, $K(1, 1)$ is a function of θ and therefore of θ . We can thus rewrite Equation (9.8) as

$$t = \frac{\sqrt{d}}{\sqrt{g}} \Phi(\theta), \quad (9.9)$$

where θ is the (dimensionless) measurement of θ , and Φ is some particular real-valued function. The usual law for the period of the pendulum using Newton's laws is basically the same as Equation (9.9) but with $\Phi(\theta) = \sin(\theta)$. Thus by using Newton's laws, $\Phi(\theta)$ can be completely determined without having to resort to experiment. However, by using the above dimensional analysis arguments, certain laws of pendulums can also be determined without having to resort to experimentation. For example, if there are two pendulums of measured lengths d_1 and d_2 (measured of course in the same unit) that are released at the same angle with the vertical, and their resulting periods are t_1 and t_2 (measured in the same unit), then it immediately follows from Equation (9.9) that

$$\frac{t_1}{t_2} = \frac{\sqrt{d_1}}{\sqrt{d_2}}. \quad (9.10)$$

Note that Equation (9.10) does not depend on which units are used to measure distance and time and that the gravitational constant is not mentioned in the statement of the result or its conclusion in the guise of Equation (9.10).

This application of dimensional analysis raises a number of questions and concerns, for example:

- (i) Why are the variables ratio scalable?
- (ii) What is the general qualitative relationship between the variables that allows for dimensional invariance to occur?
- (iii) Is the method of dimensional analysis rigorous?
- (iv) If not, can it be made to be?
- (v) Can these be generalized so that they apply outside of physics, for example to some parts of psychology or economics?

All of these questions and concerns have been investigated by measurement theorists and answered. Questions (i) and (ii) have been answered by a theory of measurement known as the “representational theory,” which is discussed in Sections 9.3 and 9.6. Questions (iii) to (v) are addressed in Part II, where a general approach to the use of scale types of individual variables combined with minimal assumptions about their interrelationships are used to derive laws interrelating them.

Notation 9.1 Throughout this chapter, we use:

- \mathbb{R} for the set of real numbers,
- \mathbb{R}_+ for $\{x \in \mathbb{R} \mid x \geq 0\}$, the set of non-negative real numbers,
- \mathbb{R}_{++} for $\{x \in \mathbb{R} \mid x > 0\}$, the set of positive real numbers,
- \iff which stands for “if and only if.”

9.3 The Representational Theory of Measurement

In Stevens' theory of measurement, any rule that assigns numbers to objects is the basis for a measurement scale. The representational theory of measurement provides a specific rule. It generalizes a key idea used in papers by the mathematician–physicist–psychologist–physiologist Helmholtz (1887) and the mathematician Hölder (1901) that provides a mathematical foundation for physical measurement. Their foundation consisted of providing axiomatic, qualitative theories for fundamental physical qualities like length, mass, time, etc., and showing through a theorem that these qualities can be represented numerically as ratio scales. Over time, there appeared several generalizations of Helmholtz's and Hölder's approach to measurement. Collectively, they are called here the “representational theory (of measurement).”

Helmholtz's and Hölder's theories were based on a special kind of measurement structure known as a “continuous extensive structure” that had simple, empirically based, physical interpretations.

Definition 9.2 The triple $\mathfrak{X} = \langle X, \precsim, \oplus \rangle$ is said to be a *continuous extensive structure* if \precsim is a binary relation on a set X , \oplus is a binary operation of X , and the following seven axioms hold. We write $x \prec y$ if $x \precsim y$ and $\neg(y \precsim x)$.

1. *Total ordering*: \precsim is a reflexive total ordering on X .
2. *Density*: For all x and z in X , if $x \prec z$ then for some y in X , $x \prec y \prec z$.
3. *Associativity*: \oplus is *associative*; that is,

$$(x \oplus y) \oplus z = x \oplus (y \oplus z)$$

for all x , y , and z in X .

4. *Monotonicity*: For all x , y , and z in X ,

$$x \precsim y \iff x \oplus z \precsim y \oplus z \iff z \oplus x \precsim z \oplus y.$$

5. *Solvability*: For all x and y in X , if $x \prec y$, then for some z in X , $x \oplus z = y$.
6. *Positivity*: $x \prec x \oplus y$ and $y \prec x \oplus y$, for all x and y in X .
7. *Continuous domain*: There exists a function f from X onto \mathbb{R}_{++} such that for all x and y in X ,

$$x \precsim y \iff f(x) \leq f(y).$$

In essence, Helmholtz (1887) showed a close variant to the following theorem, and Hölder substantially added rigor and generalized Helmholtz's result.

Theorem 9.3 Suppose $\mathfrak{X} = \langle X, \precsim, \oplus \rangle$ is a continuous extensive structure. Then the set \mathcal{S} of isomorphisms of \mathfrak{X} onto $\langle \mathbb{R}_{++}, \leq, + \rangle$ is a ratio scale.

The representational theory usually interprets the continuous extensive structure $\mathfrak{X} = \langle X, \precsim, \oplus \rangle$ as an idealized or qualitative empirical situation, where X is the set of objects to be measured and \precsim and \oplus are an observable, testable relation and

operation on X . Then the axioms for extensive structure can be tested empirically, except for continuous domain.

The definition of extensive structure and its axioms have various generalizations. In particular, Hölder assumed a weaker axiom (in mathematics called the *Archimedean axiom*) instead of continuous domain so that S became a ratio scale of isomorphisms *into* \mathbb{R}_+ instead of *onto* \mathbb{R}_+ . For some applications (e.g., see Luce, 1967, which provided a measurement theoretic foundation for probability theory), Theorem 9.3 needed to be formalized so that it applied to bounded structures $\langle X, \precsim, \oplus \rangle$ that had isomorphisms into a closed additive interval $\langle [0, 1], \leq, + \rangle$, and Luce and Marley (1969) provided an axiomatization for this. Krantz, Luce, Suppes, and Tversky (1971) and Luce, Krantz, Suppes, and Tversky (1990) replaced Solvability for the Hölder and the Luce–Marley axiomatizations with the condition of *restricted solvability*, which says that for all a and b in X , if $a \prec b$, then there exists x in X such that $a \prec x \prec b$. This makes extensive measurement systems easier to test experimentally. Falmagne (1975) (see also Falmagne, 1971) further provided extensive structures and theorems for isomorphisms from $\langle X, \precsim, \oplus \rangle$ into the open interval additive system $\langle]r, s[, \leq, + \rangle$, $0 < r < s < \infty$ and $3r < s$. This is consistent with the existence of infinitely small or infinitely large elements of X , but does not assume that such elements exist. In many applications, very small elements and very large elements have qualitatively different behaviors than a range of medium-sized elements, and this system allows the technique of extensive measurement to be applied to these medium-sized elements. A description of some of these generalizations with applications can be found in chapter 3 of Krantz *et al.* (1971). A probabilistic theory of extensive measurement is described in Falmagne (1980).

The measurement of mass by an *equal arm pan balance* is an example of the kind of continuous extensive structure that Helmholtz and Hölder had in mind. It is a measurement instrument consisting of two pans, where objects of various masses are put into one or both pans. Objects a and b are said to be “equivalent in mass,” in symbols $a \sim b$, if when placed in opposite pans, a balances b . It is assumed that \sim is an equivalence relation. Let X be the set of \sim -equivalence classes. A binary relation \precsim and an operation \oplus are defined as follows on the set X :

- $\beta \precsim \alpha$ if there exist objects x in α and y in β such that if x and y are placed in opposite pans, either they balance or the pan with x becomes lower than the one with y .
- $\alpha \oplus \beta = \gamma$ if there exist x in α , y in β , and z in γ such that when x and y are placed in the same pan and z in the opposite, the result balances.

It is assumed that, using observations about \precsim and \oplus , X passes the empirical testing of the axioms for a continuous extensive structure except for continuous domain, which instead is assumed as a theoretical condition that is necessary to have the idealized empirical structure to match the underlying physical theory of mass.

In classical physics, all the fundamental physical variables – distance, time, mass, charge, etc. – are assumed to be measurable by methods ultimately based on

Theorem 9.3 or some of its generalizations. The measurements of other physical variables, e.g., momentum, energy, density, etc., are derivable from these through various means.

Some prominent measurement theorists, e.g., Campbell (1920, 1928), concluded that fundamentally there were two kinds of measurement: a strong kind and a weak kind. He considered *extensive measurement*, which is based on extensive structures, to be the strong kind. He regarded this form of measurement as providing justifications for the use of powerful methods of mathematics such as calculus. The weak kind consisted of forms of measurement that relied on the counting of objects or the numerical ordering of objects. For Campbell, the strong kind led to mathematical science and the weak kind to a form of qualitative science. In particular, Campbell thought that measurement in psychology, if rigorously done, consisted of the weak kind. These views dominated the mathematical thinking about measurement until the 1950s. Scott and Suppes (1958) formulated a generalization of what they considered to be the key idea behind the usefulness of extensive measurement. They write:

A primary aim of measurement is to provide a means of convenient computation. Practical control or prediction of empirical phenomena requires that unified, widely applicable methods of analyzing the important relationships between the phenomena be developed. Imbedding the discovered relations in various numerical relational systems is the most important such unifying method that has yet been found.

(Scott & Suppes, 1958, pp. 116–117.)

(“Relational systems” are structures like \mathfrak{X} and \mathfrak{N} described in the next paragraph. The kind of imbeddings they employ are homomorphisms and are described later in that paragraph.)

Let X be a qualitative or empirical domain to be measured. Qualitative or empirical relations Q_j on X , $j \in J$, (including 0-ary relations, i.e., elements of X) are used to provide a measurement structure on X . This structure, which has the form $\mathfrak{X} = \langle X, Q_j \rangle_{j \in J}$, is used to measure the objects in X , which results in a numerical structure $\mathfrak{N} = \langle N, T_j \rangle_{j \in J}$, where $N \subseteq \mathbb{R}$ and the $T_j, j \in J$ are relations on N . The set X and the relations $Q_j, j \in J$, are called the *primitives* of \mathfrak{X} . It is assumed that Q_j and T_j match in the sense that Q_j is a n_j -ary relation if and only if T_j is a n_j -ary relation. Then φ is said to be a *homomorphism from \mathfrak{X} into \mathfrak{N}* if φ is a function from X into N such that for each j in J ,

- (i) if $n_j = 0$, then $\varphi(Q_j) = T_j$; and
- (ii) if $n_j > 0$, then for all x_1, \dots, x_{n_j} in X ,

$$Q_j[x_1, \dots, x_{n_j}] \iff T[\varphi(x_1), \dots, \varphi(x_{n_j})].$$

There is another usage of “homomorphism” common in mathematics that employs “if ... then” instead.

The special case where φ is a one-to-one function onto N is called an *isomorphism*. In Theorem 9.3, the extensive structure was measured by isomorphisms.

The Scott–Suppes theory contained no meaningfulness concept. Suppes and Zinnes (1963) extended the theory to include the following form of Stevens' meaningfulness concept:

Definition 9.4 Let \mathcal{S} be the set of homomorphisms from the qualitative structure \mathfrak{X} into the numerical structure \mathfrak{N} , and let $R(x_1, \dots, x_n)$ be a relation on the domain of \mathfrak{N} . Then \mathcal{S} is said to be *quantitatively \mathcal{S} -meaningful* if for each f and g in \mathcal{S} and each a_1, \dots, a_n in the domain of \mathfrak{X} ,

$$R[f(a_1), \dots, f(a_n)] \iff R[g(a_1), \dots, g(a_n)].$$

Roberts (1985) and others have applied the Suppes and Zinnes extended form of the representational theory to cover many kinds of meaningfulness issues in the behavioral sciences. The formalization of meaningfulness used by Falmagne and Narens (1983) and described in Part II, Section 9.9.4 of this chapter, for example, has much in common with quantitative \mathcal{S} -meaningfulness. The fundamental difference, however, is that in the Falmagne and Narens (1983) approach, the entities about which the invariance holds (for example, the homomorphisms f and g) are explicitly indicated in the notation for the relations, and *families* of relations are the basic object of study, rather than individual relations. This shift of focus to families of relations is a key theme of Part II of this chapter.

Pfanzagl (1968) developed a different meaningfulness concept that applies directly to the qualitative or empirical structure. It is similar in many respects to quantitative \mathcal{S} -meaningfulness. The following definition extends Pfanzagl's development to include non-representational theories, including Stevens' (1946, 1951) theory.

Let \mathfrak{X} and \mathfrak{N} be as above, \mathcal{S} be a scale family on X , and R be an n -ary relation on X . Then R is said to be *qualitatively \mathcal{S} -meaningful* if there exists an n -ary relation T , $n \geq 0$, on the domain of \mathfrak{N} such that for all φ in \mathcal{S} ,

- (i) if $n = 0$ (that is, $R \in X$), then $\varphi(R) = T$ (where T is 0-ary), and
- (ii) if $n > 0$, then for all x_1, \dots, x_n in X ,

$$R[x_1, \dots, x_n] \iff T[\varphi(x_1), \dots, \varphi(x_n)].$$

In the representational theory, there are many equally valid numerical representing structures for a given qualitative structure. As the following example shows, this clearly presents a problem for use of qualitative \mathcal{S} -meaningfulness in the homomorphism version of the theory.

Example 9.5 Let

1. $X = \{a, b, c\}$, \precsim be a total ordering on X such that $a \prec b \prec c$,
2. $\mathfrak{X} = \langle X, \prec \rangle$, $\mathfrak{M} = \langle \{1, 2, 3\}, < \rangle$, $\mathfrak{N} = \langle \mathbb{R}, < \rangle$,
3. $R(z)$ be the following relation on X : $R(z)$ if and only if $z = b$, and
4. $U(z)$ be the following relation on $\{1, 2, 3\}$: $U(z)$ if and only if $z = 2$.

Then the set \mathcal{S} of homomorphisms of \mathfrak{X} into \mathfrak{N} consists of the single function φ , where $\varphi(a) = 1$, $\varphi(b) = 2$, and $\varphi(c) = 3$. Thus R is qualitatively \mathcal{S} -meaningful, because we have $R(x) \iff U(\varphi(x))$ for each x in X .

Let \mathcal{T} be the set of homomorphisms of \mathfrak{X} into \mathfrak{M} . It will be shown by contradiction that R is not qualitatively \mathcal{T} -meaningful.

Suppose R were \mathcal{T} -meaningful. Then a relation T on \mathbb{R}_{++} can be found such that for all x in X and all ψ in \mathcal{T} ,

$$R(x) \iff T(\psi(x)). \quad (9.11)$$

Let φ be as above. Then φ is in \mathcal{T} . Thus, applying Equation (9.11) to φ and c yields,

$$R(c) \iff T(\varphi(c)) \iff T(3).$$

By the definition of “ R ”, not $R(c)$ holds. Thus

$$\text{not } T(3). \quad (9.12)$$

Let θ be the following function on X : $\theta(a) = 1$, $\theta(b) = 3$, and $\theta(c) = 4$. Then θ is in \mathcal{T} . Applying Equation (9.11) to θ and b then yields,

$$R(b) \iff T(\theta(b)) \iff T(3).$$

By the definition of “ R ”, $R(b)$ holds. Thus $T(3)$, contradicting Equation (9.12).

To avoid having the meaningfulness of a qualitative relation depend on the choice of the numerical representing structure, various measurement theorists have suggested using endomorphism or isomorphism invariance as a meaningfulness concept, as follows.

Let $\mathfrak{X} = \langle X, Q_j \rangle_{j \in J}$, where for each j in J , Q_j is an n_j -ary relation on X , $n_j \geq 0$. Then α is said to be an *endomorphism* of \mathfrak{X} if α is a function from X into \mathfrak{X} such that for all j in J ,

- (i) if $n_j = 0$, then $\alpha(Q_j) = Q_j$, and
- (ii) if $n_j > 0$, then for all x_1, \dots, x_{n_j} in X ,

$$Q[x_1, \dots, x_{n_j}] \iff Q[\alpha(x_1), \dots, \alpha(x_{n_j})].$$

An endomorphism of \mathfrak{X} that is a one-to-one function and is onto X is called a *symmetry* (or alternatively, an *automorphism*).

Let $\mathfrak{X} = \langle X, Q_j \rangle_{j \in J}$, where for each j in J , Q_j is an n_j -ary relation on X , $n_j \geq 0$. Let R be an n -ary relation on X . Then:

- R is said to be \mathfrak{X} -*symmetry meaningful* if for each symmetry α of \mathfrak{X} and each x_1, \dots, x_n in X ,

$$R[x_1, \dots, x_n] \iff R[\alpha(x_1), \dots, \alpha(x_n)].$$

- R is said to be \mathfrak{X} -*endomorphism meaningful* if for each endomorphism α of \mathfrak{X} and each x_1, \dots, x_n in X ,

$$R[x_1, \dots, x_n] \iff R[\alpha(x_1), \dots, \alpha(x_n)].$$

Theorem 9.6 Let $\mathfrak{X} = \langle X, Q_j \rangle_{j \in J}$, where for each j in J , Q_j is an n_j -ary relation on X , R is an n -ary relation on X , $n \geq 0$, and S is a scale of homomorphisms of \mathfrak{X} into the numerical structure \mathfrak{N} . Then the following statements hold. (They follow directly from the above definitions.)

1. If R is qualitatively S -meaningful, then R is endomorphism meaningful.
2. If R is endomorphism meaningful, then R is symmetry meaningful.
3. R is symmetry meaningful if and only if it is S -quantitatively meaningful and each element of S is a symmetry.

Examples can be provided that show that the converses of Statements 1 and 2 of Theorem 9.6 do not hold (e.g., Narens, 1981, page 34).

In terms of this section's terminology, Narens (1981) makes the following comment about the above qualitative meaningfulness concepts.

A number of concepts of qualitative meaningfulness have been presented, and the problem remains of deciding which, if any, is the “correct” concept. It is my view that there is no single correct concept of meaningfulness. I believe that in the final analysis the choice of the “correct” meaningfulness concept for a structure will not be determined solely by the structure, but in general will depend upon features of the intended measurement application. What we have today is a handful of successful applications of the various meaningfulness concepts; what is still lacking is a general theory of meaningfulness and inference based upon meaningfulness. The meaningfulness concepts presented above are attempts to abstract the common core of this handful of successful applications, and are not based upon any detailed philosophical analysis, and thus their usefulness and generality are somewhat in doubt. Hopefully in the future someone will find a more direct and comprehensive approach to this important problem. For purposes of basing theories of meaningfulness on the meaningfulness concepts considered, symmetry meaningfulness has the greatest applicability, mainly because the most important structures that appear in measurement have an abundance of symmetries. Endomorphism meaningfulness and qualitative S -meaningfulness, when they do not coincide with symmetry meaningfulness, thus far have had far fewer applications.

I also believe that these two latter concepts have inherent difficulties [as bases for theories of meaningfulness], which arise from the fact that representations of the qualitative structure are only required to be *into* (rather than *onto*) the [numerical] representing structure.

Interesting enough, it is this “into” property ... that make [meaningfulness based on endomorphisms and qualitative S -meaningfulness of relations] natural concepts for measurement. However, to my knowledge, the practice of using “into” representations for the general measurement case has never been philosophically justified. The situations where “into” representations have been useful are rather special.

(Narens, 1981, pp. 45–46.)

Since 1981, two additional meaningfulness concepts have appeared: a form of quantitative meaningfulness due to Falmagne and Narens (1983) that is used as part of the basis for the theory presented in Part II of this chapter, and a radically different concept that is presented in Part III.

9.4 Meaningfulness in Geometry

“Symmetry” is both an important theoretical concept and a powerful multi-faceted scientific tool. Nineteenth-century mathematicians and physicists discovered that the understanding of a mathematical or scientific situation in terms of its symmetries often provided profound insight into the underlying geometric situation. This led to the development of symmetry methods for doing mathematical inference and characterization. In 1872, the mathematician Felix Klein developed a prominent program in geometry for this (Klein, 1872). Klein’s program was based on geometrical symmetries and was designed to unify the qualitative and quantitative approaches to geometry that existed at the time. Its core ideas are very close to those of the representational theory of measurement. The main difference is that its focus was on a complicated area of mathematics – geometry – whereas the representational theory focus is on science. For example, lurking in the background of the representational theory are concerns, e.g., empiricism and error, that are not relevant for geometry.

In mathematics, three approaches to geometry have evolved. The first is a *synthetic* approach. It was developed by ancient geometers, and is the content of Euclid’s famous third-century BC treatise. This axiomatic and deductive approach of Euclid is today called “plane Euclidean geometry.” Variants of the approach were also systematically applied to other geometries by nineteenth-century geometers. These became known as “synthetic approaches to geometries,” because they were based on formalized methods having intuitive geometric content and significance. Their formalized methods have evolved into what today is called “the axiomatic method,” which has been applied to many non-geometric domains.

The second is the *analytic* approach to geometry. It began as a coordinate and algebraic approach that was developed by Descartes in the early seventeenth century. Its current version takes n -tuples of real numbers as coordinates and considers geometric objects as functions and sets of equations involving the coordinates. Mathematical operations, including those from analysis and algebra, are allowed to be performed on them in order to solve geometric problems. Such operations are not required to have intuitive geometric content or significance: *Only the statement of the geometric problem and its solution are required to have intuitive geometric content or significance.* Analytic geometric methods often permit much simpler solutions to geometric problems than synthetic ones. Both synthetic and analytic approaches to geometry flourished in nineteenth-century mathematics. New geometries were discovered, and both synthetic and analytic approaches to them were developed. Sometimes axiomatic systems for the same geometry were based on different geometric concepts. Such developments raise two important theoretical issues:

- (1) When do different formulations describe “the same geometry”?
- (2) What does it mean for concepts and relations of a geometry to be appropriate concepts and relations in “the same geometry” formulated in terms of different primitive relations?

Answers to these questions ultimately lead to a systematic classification of geometries and provided for a way to bridge the synthetic and analytic approaches. Klein (1872) proposed answers in what has become known as the *Erlanger program*. The answers are based on transformation groups and their invariants.

Convention 1 Throughout this section, the following notation and conventions are used:

- $\mathfrak{X} = \langle X, R_j \rangle_{j \in J}$ is a qualitative structure;
- G is the set of symmetries of \mathfrak{X} ;
- $\mathfrak{N} = \langle N, S_j \rangle_{j \in J}$ is a representing numerical structure, where N is a set of n -tuples of real numbers;
- \mathcal{S} is a scale family of isomorphisms from \mathfrak{X} onto \mathfrak{N} ;
- $\varphi \in \mathcal{S}$.

It is immediate that for each symmetry α of \mathfrak{X} , $\varphi \circ \alpha$ is an isomorphism of \mathfrak{X} onto \mathfrak{N} (because a composition of isomorphisms is again an isomorphism), and therefore $\varphi \circ \alpha$ is in \mathcal{S} . It is also immediate that for each ψ in \mathcal{S} ,

$$\psi^{-1} \circ \varphi = \beta$$

is a symmetry of \mathfrak{X} (again because a composition of isomorphisms is an isomorphism). Thus, because

$$\varphi = \psi \circ \beta,$$

it follows that

$$\varphi \circ \beta^{-1} = \psi,$$

and thus

$$\alpha \mapsto \varphi \circ \alpha$$

defines a one-to-one function from the set G of symmetries of \mathfrak{X} onto the scale family \mathcal{S} .

The set G of symmetries of \mathfrak{X} together with the operation of functional composition \circ form a much-investigated, important mathematical structure called a “transformation group.” Formally, $\mathcal{H} = \langle H, \circ \rangle$ is said to be a *group of transformations* (or *transformation group*) if for some non-empty set Y ,

- (i) the identity function ι_Y on Y is in H ;
- (ii) each element of H is a one-to-one function from Y onto Y ; and
- (iii) for all f and g in H , $f \circ g^{-1}$ is in H .

By convention, for a group $\mathcal{H} = \langle H, \circ \rangle$, “ $f \in \mathcal{H}$ ” stands for “ $f \in H$,” and “ H is a group” for “ \mathcal{H} is a group.”

Convention 2 Throughout the remainder of this section let

$$\mathcal{G} = \langle G, \circ \rangle.$$

We call \mathcal{G} the *symmetry group* of \mathfrak{X} (or *transformation group* of \mathfrak{X}).

The following theorem is immediate from the definitions of symmetry and isomorphism.

Theorem 9.7 *\mathcal{G} is a group of transformations.*

Let $\langle H, \circ \rangle$ be a symmetry group on Y and R an n -ary relation, $n \geq 1$, on Y . Then R is said to be an *invariant of H* if for all y_1, \dots, y_n in Y and all h in H ,

$$R(y_1, \dots, y_n) \iff R[h(y_1), \dots, h(y_n)].$$

This concept of “invariant” extends to elements y of Y as follows: y is an *invariant of H* if $h(y) = y$ for all h in H . It also extends to higher-order relations on Y . For example, for each non-empty subset Z of Y , that is, for each 1-ary relation Z of Y , and each h in H , let

$$h(Z) = \{h(z) | z \in Z\}.$$

Then Z is invariant under H if $h(Z) = Z$. Let \mathcal{W} be a non-empty set of non-empty sets of Y . Note that \mathcal{W} is an example of a particular kind of higher-order relation on Y , namely a 1-ary relation of 1-ary relations of Y . By definition, for each h in H let

$$h(\mathcal{W}) = \{h(Z) | Z \in \mathcal{W}\}.$$

Then \mathcal{W} is said to be *invariant under H* if for each h in H ,

$$h(\mathcal{W}) = \mathcal{W}.$$

As an example, consider Y to be the set of points of the Euclidean plane. Then each line in the plane is a non-empty subset of Y . Let \mathcal{L} be the set of lines. Then \mathcal{L} corresponds to the Euclidean concept of “line.” It is a non-empty set of non-empty subsets of Y . Let H be the group of rotations of points of Y about a particular point of Y . Then elements of H take lines into lines, that is, for each h in H ,

$$h(\mathcal{L}) = \mathcal{L},$$

that is, \mathcal{L} is a “higher-order invariant under H .”

Higher-order invariants extend to relations as well. The basic idea is the same as for higher-order sets: If R is an n -ary relation and $R(S_1, \dots, S_n)$ holds, then by methods described later in Part III, $h(S_1), \dots, h(S_n)$ are defined. Thus, letting by definition, $h[R(S_1, \dots, S_n)] = R[h(S_1), \dots, h(S_n)]$, it follows that R is invariant under h if and only if for all S_1, \dots, S_n in the domain of R ,

$$R(S_1, \dots, S_n) \iff R[h(S_1), \dots, h(S_n)] = h[R(S_1, \dots, S_n)],$$

that is, $h(R) = R$.

Higher-order relations and invariants appear throughout mathematics and science. Note that representational theory was formulated in terms of first-order relations – relations or elements of a non-empty set – and its (symmetry) meaningfulness concept was about first-order invariants.

The Erlanger program identifies geometries with transformation groups: *Two geometric structures (either synthetic or analytic) are said to specify the same geometry if they have isomorphic transformation groups.* Synthetic geometries consist of a structure of primitives,

$$\mathfrak{Y} = \langle Y, T_j \rangle_{j \in J},$$

where for each j in J , T_j is either an element of Y (a 0-ary relation), or an n_j -ary relation on Y , $n_j \geq 1$, or a higher-order relation on Y (which includes a relation of m -ary relations, $m \geq 0$, on Y , a relation of relations of m -ary relations on Y , etc.), and axioms about the primitives of \mathfrak{Y} . The transformation group $\langle H, \circ \rangle$ of \mathfrak{Y} consists of the set H of transformations on Y that leave \mathfrak{Y} 's primitives invariant. (It is immediate that $\langle H, \circ \rangle$ is a group.) Axioms about the primitives are used to derive group-theoretic properties that specify H up to an isomorphism.³

The third approach to geometry is a transformational one based on the Erlanger program: *A geometry \mathfrak{Y} is a transformation group $\langle H, \circ \rangle$ on a non-empty set Y . The relations – 0-ary, n -ary, higher-order – that are invariant under H belong to \mathfrak{Y} ; those that are not invariant do not belong to \mathfrak{Y} .*

As an example, consider the set \mathcal{E} of points consisting of ordered pairs of real numbers. In terms of Cartesian coordinates, the Euclidean distance function d between elements (x, y) and (u, v) of \mathcal{E} is defined by

$$d[(x, y), (u, v)] = \sqrt{(x - u)^2 + (y - v)^2},$$

with an orientation for angles specified so that counterclockwise from the positive abscissa to the positive ordinate determines positive orientation. The function d and this orientation specify an analytic geometry on \mathcal{E} . The set E of transformations that leave d and the orientation invariant are known to be those generated by translations and rotations. Thus $\langle E, \circ \rangle$ provides a transformational description **E** of a geometry on \mathcal{E} , traditionally called “Euclidean plane geometry.” There are a number of synthetic axiomatizations of **E**, sometimes with differing primitives. The symmetry group for the structure of primitives of each of these axiomatizations is isomorphic to $\langle E, \circ \rangle$.

9.5 Comparison of Geometric and Measurement-theoretic Concepts

The scientific structures dealt with by most of this chapter are based on open (often unbounded) intervals of the reals. These in mathematics are called “continua” and they were qualitatively described by Cantor (1895) as follows:

$\langle X, \preceq \rangle$ is said to be a *continuum* if the following four statements hold:

1. *Total ordering*: \preceq is a total ordering on X .
2. *Unboundedness*: $\langle X, \preceq \rangle$ has no \preceq -greatest or \preceq -least element.

3 That is, any two structures satisfying the axioms will have isomorphic symmetry groups.

3. *Denumerable density*: There exists a denumerable subset Y of X such that for each x and z in X , if $x \prec z$ then there exists y in Y such that $x \prec y$ and $y \prec z$.
4. *Dedekind completeness*: $\langle X, \preceq \rangle$ is Dedekind complete (i.e., each non-empty \preceq -bounded subset of X has a least upper bound).

$\mathfrak{X} = \langle X, \preceq, R_j \rangle_{j \in J}$ is said to be a *continuous structure* if $\langle X, \preceq \rangle$ is a continuum and for each j , R_j is an element of X , a subset of X , or an n -ary relation on X for some n .

Cantor (1895) contains the following theorem. (See also theorem 2.2.2 of Narens, 1985, for a proof.)

Theorem 9.8 (Cantor's continuum theorem) $\mathfrak{X} = \langle X, \preceq \rangle$ is a continuum if and only if \mathfrak{X} is isomorphic to $\langle \mathbb{R}^+, \leq \rangle$.

Let $\mathfrak{X} = \langle X, \preceq, R_j \rangle_{j \in J}$ be a continuous structure. Then there exists a numerical structure called a *numerical representing structure for \mathfrak{X}* , $\mathfrak{N} = \langle N, \leq, S_j \rangle_{j \in J}$, that is isomorphic to \mathfrak{X} and is such that N is an open interval of reals (possibly infinite). (This follows from Cantor's theorem by having an isomorphism f of $\langle X, \preceq \rangle$ onto $\langle \mathbb{R}^+, \leq \rangle$ followed by an isomorphism g from $\langle \mathbb{R}^+, \leq \rangle$ onto $\langle N, \leq \rangle$ and letting $S_j = g[f(R_j)]$.)

Note by the above definitions, a continuous extensive structure defined previously in 9.2 is a continuous structure with numerical representing structure $\langle \mathbb{R}^+, \leq, + \rangle$.

With these definitions and results, the following theorem can be established showing the equivalence of previously defined meaningfulness and invariance concepts for continuous structures. The proof is straightforward and is left to the reader.

Theorem 9.9 Suppose that $\mathfrak{X} = \langle X, \preceq, R_j \rangle_{j \in J}$ is a continuous structure, G is the set of symmetries of \mathfrak{X} , and R is an n -ary relation on X . Then the following three statements are equivalent:

1. R is \mathcal{S} -quantitatively meaningful (Definition 9.4) for some scale family of isomorphisms \mathcal{S} onto a numerical representing structure for \mathfrak{X} .
2. R is invariant under G .
3. For each scale family of isomorphisms \mathcal{S}^* from \mathfrak{X} onto a numerical representing structure, R is \mathcal{S}^* -representationally meaningful (Definition 9.4).

The following shows a correspondence between some of the basic concepts of the representational theory of measurement based on a scale of isomorphisms onto a numerical structure and the Erlanger program for geometry:

qualitative structure \longleftrightarrow synthetic geometry

numerical structure \longleftrightarrow analytical geometrical structure

measurement scale isomorphism \longleftrightarrow $\left\{ \begin{array}{l} \text{isomorphism from a synthetic} \\ \text{geometry to an analytic geometry} \end{array} \right\}$

symmetry meaningful \longleftrightarrow $\left\{ \begin{array}{l} \text{belonging to the geometry} \\ \text{under consideration} \end{array} \right\}$

9.6 Dimensional Analysis within the Representational Theory

In Volume 1 of *Foundations of Measurement*, Krantz et al. (1971) applied extensive measurement to a wide range of psychological phenomena. This was accomplished by using the primitives of the qualitative structure \mathfrak{X} to *define* an appropriate version of an extensive structure \mathfrak{E} , using \mathfrak{E} to define a scale \mathcal{S} of measuring functions into a numerical structure $\langle N, \leq, + \rangle$ (or in some cases, $\langle N, \leq, \cdot \rangle$), and then using \mathcal{S} to define a numerical structure \mathfrak{N} with domain N such that \mathcal{S} is a homomorphism of \mathfrak{X} into \mathfrak{N} . Note the difference of this approach to that of Campbell and others who thought that extensive measurement was the only avenue to strong forms of measurement: they required a qualitative or empirical extensive structure $\mathfrak{Y} = \langle Y, \prec, \oplus \rangle$ such that all the primitives of \mathfrak{Y} were *observable*; Krantz et al. instead required an empirical structure \mathfrak{X} such that

- (i) all the primitives of \mathfrak{X} are observable, and
- (ii) the primitives of the extensive structure \mathfrak{E} are *definable* in terms of the primitives of \mathfrak{X} .

The method of Krantz et al. allows for a much wider application of measurement based on extensive structures than that of Campbell's.

Krantz et al. also developed measurement techniques to provide a representational measurement approach for a foundation of physical dimensional analysis.

Narens and Luce (1976) discovered that much of the program carried out by Krantz et al. could be generalized by dropping the associative axiom of extensive measurement. This provided a potentially enormously wider range for the application of the representational approach. They also formulated a key concept relating different measurement dimensions called *distributive triples* that greatly simplified the construction of the structure of physical units described in Krantz et al. This led Luce (1978) to reformulate the foundation of dimensional analysis of Krantz et al. through the use of distributive triples. In his reformulation, he showed that the key inference rule in physical dimensional analysis – called *dimensional invariance* – was logically equivalent to his distributive triples-based structure of physical units being invariant under the symmetries of the resulting structure of physical units. That is, Luce showed that dimensional analysis could be considered as a special case of the Erlanger program. Luce and Narens realized that what was important in many measurement theory applications was obtaining a scale of functions into the real numbers for the domain. The exact form of the primitives did not matter. That is, the important consideration was the transformation group of the measurement structure. This is much in line with Klein's Erlanger approach to geometry. Luce and Narens, together with Luce's student, Michael Cohen, ended up developing a general theory of measurement structures through invariance considerations. This work is summarized in Luce and Narens (1985).

We now turn to Part II of this chapter, which details some applications of the meaningfulness concept for a purpose similar to that of dimensional analysis,

namely, to uncover the possible forms of functions relating the measurement variables in a scientific context.

Part II: Meaningfulness and Abstract Axiomatization of Scientific Laws

9.7 Meaningfulness as a Principle of Theory Construction

As alluded to in Section 9.1, the work of Duncan Luce, and in particular his “*principle of theory construction*” (Luce, 1959), has inspired a great deal of work in the study of invariance in science, including the results we detail in Part II. Luce had in mind a possible linkage of the “admissible transformations” of the variables in a scientific equation to the possible form of the scientific law formalized by that equation, with “admissible transformations” meaning “changes of units.” The paper (Luce, 1959) suggested a new approach to the formal analysis of scientific laws from a foundational standpoint. The paper drew objection, though, from Rozeboom (1962a) (see also Luce, 1962; Rozeboom, 1962b), who argued that Luce’s principle was ambiguous and susceptible to several interpretations, some of them disputable.⁴

Luce (1964) attempted to give a mathematically clearer interpretation of the principle, as follows. Let x_1, x_2, \dots, x_{n+1} be $n + 1$ real variables, and let \mathcal{F}_i be the set of admissible transformations for the i th variable, $i = 1, 2, \dots, n+1$. A possible scientific law would be given by the unknown function

$$u : \prod_{i=1}^n R_i \longrightarrow R_{n+1} : (x_1, x_2, \dots, x_n) \mapsto x_{n+1},$$

where the R_i , for $i = 1, 2, \dots, n + 1$, are suitable subsets of the reals. The problem is to find the general form of the function u knowing the admissible transformation sets \mathcal{F}_i , i.e., to find the general form of the function u such that

$$u(x_1, x_2, \dots, x_n) = x_{n+1}.$$

The sets \mathcal{F}_i of admissible transformations would be known from the scale types of the variables based on the measurement situation. For example, if the i th variable were ratio-scaled, then \mathcal{F}_i would consist of all functions

$$T_i : R_i \rightarrow R_i : x \mapsto r_i x \quad (r_i \in \mathbb{R}_{++}).$$

If the i th variable were interval-scaled, then \mathcal{F}_i would consist of all functions

$$T_i : R_i \rightarrow R_i : x \mapsto r_i x + p_i \quad (r_i \in \mathbb{R}_{++}, p_i \in \mathbb{R}).$$

⁴ The introduction of this section follows closely Falmagne and Doble (2016). The rest of the chapter from this point on relies on results from Falmagne (2004, 2014, 2015) and Falmagne and Doble (2015). See also Falmagne and Narens (1983).

Or it could instead be that the i th variable had some other scale type. For more information on scale types, see Section 9.1 of this chapter, and also Krantz *et al.* (1971) or Roberts (1979).

The crux of Luce's principle (Luce 1959, 1964) is that, in the absence of "dimensional constants," admissible transformations of the independent variables should lead to an admissible transformation of the dependent variable. This is formalized in Luce (1964) as the statement that, given $T_i \in \mathcal{F}_i$, $i = 1, 2, \dots, n$, there is a transformation $D(T_1, T_2, \dots, T_n)$ in \mathcal{F}_{n+1} such that

$$u[T_1(x_1), T_2(x_2), \dots, T_n(x_n)] = D(T_1, T_2, \dots, T_n)[u(x_1, x_2, \dots, x_n)]. \quad (9.13)$$

If, for example, all the independent variables x_1, x_2, \dots, x_n are ratio-scaled and the dependent variable x_{n+1} is also ratio-scaled, then $T_i(x_i) = \alpha_i x_i$, $\alpha_i > 0$ ($1 \leq i \leq n$), and Equation (9.13) leads to the functional equation

$$u(\alpha_1 x_1, \alpha_2 x_2, \dots, \alpha_n x_n) = R(\alpha_1, \alpha_2, \dots, \alpha_n) u(x_1, x_2, \dots, x_n), \quad (9.14)$$

with $R(\alpha_1, \alpha_2, \dots, \alpha_n) > 0$, a constant (see, e.g., Aczél, Roberts, & Rosenbaum, 1986, pp. 389–390).

Equation (9.14) and similar equations were studied by Luce (1964), Osborne (1970), Aczél *et al.* (1986), and Kim (1990) under various hypotheses concerning the function u and the scale types – ratio, interval, ordinal, or log-interval – of the independent and dependent variables.

The program described in Part II, though very much inspired by this approach, departs from it in the acceptance of Equation (9.13) as an interpretation of Luce's principle. As described in Falmagne and Doble (2016, with footnotes included):

In the usual way science is conducted, the units of the measurement scales are not part of the formal notation and marked in the equations. They may be indicated in a parenthetical comment, or they may not. The Pythagorean Theorem or Newton's Law of Universal Gravitation are always stated without any mention of the units.^a This is sensible because the units of the scales have no representation in nature.^b So, any scientific equation whose form would fundamentally depend upon the units of the variables would be a poor representation of a phenomenon. But this leads us to the following interpretation of Luce's Principle:

The mathematical form of an equation formalizing a scientific law should be invariant under changes of the units of its variables;

or in the jargon of this paper: The equation should be meaningful.

Eqs. (9.13) and (9.14), while not obviously at variance with this 'meaningfulness' principle, are not faithful representations of it: nothing in the notation indicates that $T_i(x_i)$ represents a change of unit compared to x_i .

(Falmagne & Doble, 2016, p. 60.)

a In the Law of Gravity, the constant G in $F = G \frac{m_1 m_2}{r^2}$ stands for units, but the units themselves are rarely mentioned.

b The only exception is the counting measure, in the Avogadro number.

The assertion that changes of unit be made obvious through the notation is fundamental to the work presented here in Part II. For instance, the reader can compare Equation (9.14) with Equation (9.15) in the next subsection, or with the more general Equation (9.61) in Section 9.9. Following Falmagne and Narens (1983), Narens (2002a), Doble (2002), Falmagne (2004, 2014), Falmagne and Doble (2016), and Falmagne and Doble (2015), we define a meaningfulness condition in Part II in which the units of the variables are explicitly represented in the notation. The resulting heavy notation is only temporary, as we can return to the standard notation once the results of the meaningfulness assumption have been obtained.

Once suitably axiomatized, the meaningfulness condition is constraining enough that the addition of only relatively weak, intuitive assumptions – we call them “*abstract axioms*” – may suffice to obtain the exact mathematical form of a scientific or geometric law.⁵ Part II of this chapter gives seven examples of such derivations,⁶ leading to a wide variety of scientific or geometric laws.

As a final comment before turning to an example of a meaningfulness condition (for which changes of units are expressed by the notation), we note that the results in Part II focus on ratio scales, rather than interval or log-interval scales. In particular, we restrict consideration to functions of ratio-scaled variables producing ratio-scaled values. There is a sense in which this is natural, as most of the variables of physics or geometry, such as length, mass, pressure, time, etc., are expressed in terms of ratio scales. In fact, it may be that the study of invariance axioms such as those examined in this chapter will help in understanding the predominance of ratio scales, and work is underway in this regard. See Falmagne and Doble (2016, p. 60) for related thoughts, and the comments following Examples 9.57.

9.7.1 An Example of a Meaningfulness Condition

Suppose that each of $F_\alpha(x, y)$, $F_\beta(x, y)$, and $F_1(x, y)$ computes, with a different unit indicated by α , β , and 1, the length of the hypotenuse of a right triangle from the lengths x and y of the two sides of the right angle. Then we have

$$\frac{1}{\alpha}F_\alpha(\alpha x, \alpha y) = \frac{1}{\beta}F_\beta(\beta x, \beta y) = F_1(x, y) \quad (\forall x, y, \alpha, \beta > 0). \quad (9.15)$$

The expression $F_1(x, y)$, which could be rewritten $F(x, y)$, gives the length of the hypotenuse in the *initial unit*, which by convention could be the *meter*. If $x = 3$ and $y = 4$, for example, we have

$$F_1(x, y) = F(3, 4) = \sqrt{3^2 + 4^2} = 5 \quad (\text{meters}).$$

Suppose that the indices α and β indicate that decimeters and centimeters are used, respectively, in the computation of F_α and F_β . So $F_\alpha(3, 4)$ is the length of

⁵ In some cases, possibly up to the values of some numerical parameters.

⁶ We only give a couple of exemplary proofs of the results. For most of the proofs, see Falmagne and Doble (2015). For standard functional equations results, see Aczél (1966).

the hypotenuse of a right triangle with respective side lengths of 3 decimeters and 4 decimeters. We thus have $\alpha = 10$ and $\beta = 100$.

From (9.15), we obtain:

$$F_{10}(3, 4) = 10F_1\left(\frac{3}{10}, \frac{4}{10}\right) = 10\sqrt{\left(\frac{3}{10}\right)^2 + \left(\frac{4}{10}\right)^2} = 5 \quad (\text{decimeters}),$$

$$F_{100}(3, 4) = 100\sqrt{\left(\frac{3}{100}\right)^2 + \left(\frac{4}{100}\right)^2} = 5 \quad (\text{centimeters})$$

by a similar derivation. Equation (9.15) is a simple case of the meaningfulness condition. The general condition is in Definition 9.27.

As mentioned above, such a meaningfulness condition, when paired with some general abstract axioms constraining the function representing the scientific phenomenon, may yield the possible forms of the law, or a short list of forms, up to some real-valued parameters.

An example of an abstract axiom is the so-called *associativity equation* (cf. Aczél, 1966, p. 253)

$$F(F(x, y), z) = F(x, F(y, z)) \quad (9.16)$$

which is relevant to the Pythagorean Theorem. Indeed, it can be shown by a simple geometric argument (see Section 9.7.3) that the hypotenuse of a right triangle is, as a function of the two sides of the right angle, an associative function.

9.7.2 Examples of Abstract Axioms with their Representations

Here are seven examples of abstract axioms, including the associativity equation. In each case, we suppose that the functions F , G , or L are real-valued, strictly monotonic and continuous in their two ratio scale variables, which range in non-negative real intervals.

$$\text{Associativity:} \quad F(F(x, y), z) = F(x, F(y, z)). \quad (9.17)$$

$$\text{Permutability:} \quad F(F(x, y), z) = F(F(x, z), y). \quad (9.18)$$

$$\text{Quasi-permutability:} \quad F(G(x, y), z) = F(G(x, z), y). \quad (9.19)$$

$$\text{The translation equation:} \quad F(F(x, y), z) = F(x, y + z). \quad (9.20)$$

$$\text{Bisymmetry:} \quad F(F(x, y), F(z, w)) = F(F(x, z), F(y, w)). \quad (9.21)$$

$$\text{Autodistributivity:} \quad \begin{cases} F(x, F(y, z)) = F(F(x, y), F(x, z)) \\ F(F(x, y), z) = F(F(x, z), F(y, z)). \end{cases} \quad (9.22)$$

*The abstract Lorentz–FitzGerald–Doppler axiom or LFD axiom:*⁷

$$L(L(\ell, v), w) = L(\ell, v \oplus w), \quad (9.23)$$

where $L(\ell, v)$ may⁸ denote the length of a rod measured by an observer moving at the speed v with respect to the rod, on a line parallel to the rod (with ℓ and L being measured in the same units), and \oplus stands then for the (abstract) addition of velocities in special relativity. (The exact definition of these conditions is given in Section 9.8.)

The representational consequences of Equations (9.17)–(9.22) for the functions F and G have been worked out in the functional equations literature; see Aczél (1966, chapter 6) and Ng (2016, chapter 3). Under suitable background conditions (e.g. monotonicity/continuity/solvability), there exist functions f , g and m , and in the case of Equations (9.28) and (9.29), a number $q \neq 0, 1$, such that the following representations hold.

The associativity equation implies:

$$F(x, y) = f(f^{-1}(x) + f^{-1}(y)). \quad (9.24)$$

The quasi-permutability equation implies:

$$F(x, y) = m(f(x) + g(y)), \quad (9.25)$$

with

$$G(x, y) = f^{-1}(f(x) + g(y)). \quad (9.26)$$

The translation equation implies:

$$F(x, y) = f(f^{-1}(x) + y). \quad (9.27)$$

The bisymmetry equation implies:

$$F(x, y) = f\left((1 - q)f^{-1}(x) + qf^{-1}(y)\right). \quad (9.28)$$

The autodistributivity equations imply:

$$F(x, y) = f\left((1 - q)f^{-1}(x) + qf^{-1}(y)\right). \quad (9.29)$$

Furthermore, Falmagne and Doignon (2010) proved that, under some background conditions on the pair (L, \oplus) of the abstract Lorentz–FitzGerald–Doppler equation (9.23), there exists some function u and some real parameters c and ξ , such that:

$$L(\ell, v) = \ell \left(\frac{c - u(v)}{c + u(v)} \right)^{\xi} \quad (9.30)$$

$$v \oplus w = u^{-1} \left(\frac{u(v) + u(w)}{1 + \frac{u(v)u(w)}{c^2}} \right) \quad (9.31)$$

⁷ Equation (9.23) is consistent with both the Lorentz–FitzGerald contractions equation (9.32) and the Doppler effect equation (9.56).

⁸ These quantities have different interpretations in the case of the Doppler effect.

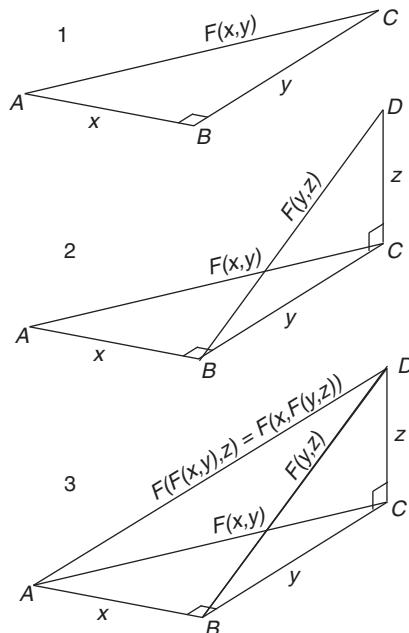
(see Lemma 9.26). We refer to each of the Equations (9.24)–(9.29) and the pair (9.30)–(9.31) as an *abstract representation* of the corresponding abstract axiom. The exact results leading to these abstract representations are stated in various lemmas of Section 9.8.

In Section 9.11, we pair a meaningfulness condition (defined in Section 9.9) with these abstract representations and find that the representations can be transformed into scientific or geometric laws, possibly defined up to some parameters. For example, the abstract representation of the associativity equation, Equation (9.24), may be transformed into the equation of the Pythagorean theorem, specified up to the exponent parameter. We now present an informal argument of this.

9.7.3 Associativity and the Pythagorean Theorem

Right triangle geometry provides a fitting context in which the associativity axiom arises. In particular, if $F(x, y)$ denotes the length of the hypotenuse of a right triangle with side lengths x and y , then the associativity axiom must be satisfied. When paired with the meaningfulness condition, this implies that the Pythagorean Theorem must hold, up to the exponent, for the family of functions measuring the hypotenuse (see Lemma 9.14 and Theorem 9.45). To see that associativity holds for the function F , consider the argument shown in the figure.

We see that the length of the hypotenuse of a right triangle is an associative function of the lengths of the two sides.



1. Draw the right triangle $\triangle ABC$ with hypotenuse \overline{AC} of length $F(x, y) = AC$. Its side lengths are $x = AB$ and $y = BC$.
2. Draw the segment \overline{CD} of length $z = CD$, perpendicular to the plane of the points A , B , and C . Draw the segment \overline{BD} . As $\triangle BCD$ is a right triangle by construction, the length of its hypotenuse is $F(y, z) = BD$.
3. Draw the segment \overline{AD} . Both $\triangle ABD$ and $\triangle ACD$ are right triangles: \overline{CD} is perpendicular to the plane of $\triangle ABC$, and \overline{AB} is perpendicular to the plane of $\triangle BCD$. The common hypotenuse of $\triangle ABD$ and $\triangle ACD$ is \overline{AD} . Its length is $F(F(x, y), z) = F(x, F(y, z))$.

Figure 9.1 The associativity of the hypotenuse of a right triangle.

In Theorem 9.45, it is shown that, assuming the associativity axiom holds for a homogeneous⁹ function F , there is only one possible form for F :

$$F(x, y) = (x^\theta + y^\theta)^{\frac{1}{\theta}} \quad (\text{for some parameter } \theta > 0).$$

Furthermore, as shown in Remark 9.46, if a couple of reasonable postulates of geometry are assumed, then we must have $\theta = 2$. So, this gives another proof of the Pythagorean theorem.

9.7.4 Order-invariance under Transformations

Another tack undertaken for studying invariance via meaningfulness is described in Section 9.12. This approach, due to Falmagne (2004), concerns families or classes of transformations of the input variables that do not alter the order of the output variable (the dependent variable).

As overviewed in Sections 9.1 through 9.6 of this chapter (see also Suppes, 2002), the study of structure invariance under a set of transformations has a long and distinguished history in science and mathematics. As discussed in Section 9.12, the fact that not all families or classes of transformations are meaningful¹⁰ sets restrictions on the possible transformations that leave the output variable order-invariant. The combination of meaningfulness axioms and this type of order-invariance can go a long way toward determining the functional form of a scientific law.

For an example of the usefulness of this order-invariance, consider the function L of the *Lorentz–FitzGerald contraction*, or *LF-contraction* for short:

$$[\text{LF}] \quad L(\lambda, v) = \ell \sqrt{1 - \left(\frac{v}{c}\right)^2}. \quad (9.32)$$

Note that this is a special case of the function L in Equation (9.23). Suppose that L also satisfies the following order-invariance: for all non-negative ℓ, v, ℓ' and v' we have, for some strictly increasing function f ,

$$L(\ell, v) \leq L(\ell', v') \iff L(f(\ell), v) \leq L(f(\ell'), v'). \quad (9.33)$$

As detailed in Section 9.12, if both Equation (9.32) and Condition (9.33) are satisfied, then the form of the function f is limited to $f(\ell) = \lambda\ell$ for some positive real number λ . This may be surprising, as there are many functions f satisfying Condition (9.33).

In the rest of Part II of this chapter, we give a number of basic definitions and results leading to the axiomatization of several important cases of scientific or mathematical laws. We return several times to the examples of the Lorentz–FitzGerald contraction, Beer’s law, the Pythagorean theorem, and the monomial

⁹ A function $F : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ is *homogeneous* if $\alpha F(x, y) = F(\alpha x, \alpha y)$ for all $\alpha, x, y > 0$.

¹⁰ That is, meaningful for a given collection of scientific functions; see Definitions 9.74 and 9.75.

laws. We mostly consider what we call *self-transforming* collections of scientific functions (see Definition 9.28), that is, functions having the scale of their output identical to the scale of their first variable. (The monomial laws, which we discuss in Subsection 9.12.7, are an exception; these are not necessarily self-transforming.) As mentioned, we focus on functions of two ratio scale variables. There is a good deal of room for elaborating the theory presented, but it is hoped that the basic concepts given and the results shown will inform the reader of some general approaches.

9.7.5 Some Basic Concepts

Definition 9.10 We recall that \mathbb{R}_{++} stands for the positive real numbers, and \mathbb{R}_+ for the non-negative real numbers. For some positive integer $n \geq 2$, let J_1, J_2, \dots, J_{n+1} be $n + 1$ non-negative, real intervals of possibly infinite length. An *n-dimensional (numerical) code*, or an *n-code* for short, is a function

$$F : J_1 \times \dots \times J_n \longrightarrow J_{n+1} \quad (9.34)$$

that is strictly monotonic and continuous in its n arguments, and strictly increasing in its first argument. As we mostly deal with 2-codes, we sometimes simplify our language and just write “code” to mean “2-code.”

Definition 9.11 Let J, J' and H be non-negative, real intervals of possibly infinite length. A code $F : J \times J' \rightarrow H$ is *solvable* if it satisfies the following two conditions.

- [S1] If $F(x, t) < p \in H$, there exists $w \in J$ such that $F(w, t) = p$.
- [S2] The function F is *1-point right solvable*, that is, there exists a point $x_0 \in J$ such that for every $p \in H$, there is a $v \in J'$ satisfying $F(x_0, v) = p$. In such a case, we may say that F is x_0 -solvable.

By the strict monotonicity of F , the points w and v of [S1] and [S2] are unique.

A code $F : J \times J' \longrightarrow H$ is *reducible on both sides* if $F(x, y) = F(x, z)$ or $F(y, x) = F(z, x)$ only if $y = z$ (Aczél, 1966, p. 255).

A code $F : J \times J \rightarrow H$ is *symmetric* if $F(x, y) = F(y, x)$ for all $x, y \in J$.

A code $F : \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ is *homogeneous*, or satisfies *homogeneity*, if $F(\gamma x, \gamma y) = \gamma F(x, y)$ for all $x, y, \gamma \in \mathbb{R}_{++}$.

Definition 9.12 Let J, J' and H be as described in Definition 9.11 above, and let H' be a non-negative, real interval of positive, possibly infinite length. Two functions $F : J \times J' \xrightarrow{\text{onto}} H$ and $G : J \times J' \xrightarrow{\text{onto}} H'$ are *comonotonic* if

$$F(x, s) \leq F(y, t) \iff G(x, s) \leq G(y, t), \quad (x, y \in J; s, t \in J'). \quad (9.35)$$

In such a case, the equation

$$M(F(x, s)) = G(x, s) \quad (x \in J; s \in J') \quad (9.36)$$

defines a strictly increasing, continuous function $M : H \xrightarrow{\text{onto}} H'$. Note that the comonotonicity relation is transitive.

9.8 Representation Theorems of the Seven Abstract Axioms

Here we formally restate the results involving (9.24) through (9.29).

9.8.1 The Associativity Equation

Definition 9.13 Let J be an open interval, possibly infinite. A code $F : J \times J \rightarrow J$ satisfies the *associativity equation*, or equivalently, is *associative* if, for all x, y and z in J we have

$$F(F(x, y), z) = F(x, F(y, z)). \quad (9.37)$$

As we saw in Subsection 9.7.3, the length of a right triangle's hypotenuse is an associative function of the lengths of the two sides. Many other functions satisfy associativity as well, which often applies to binary operations. However, we shall see in Theorem 9.45 that for a meaningful collection of self-transforming codes, and assuming associativity and homogeneity, only one representation is possible for the collection. This stems in part from the following abstract representation for the associativity equation.

Lemma 9.14 *Let J be an open interval, and suppose that some function $F : J \times J \rightarrow J$ is reducible on both sides. Then the function F is associative if and only if, for some continuous, strictly monotonic function f , we have*

$$F(x, y) = f(f^{-1}(x) + f^{-1}(y)). \quad (9.38)$$

(For a proof, see Aczél, 1966, p. 256.) Note that if the function F is strictly increasing in both variables, then it must be reducible on both sides.

9.8.2 Permutability and Quasi-permutability

Definition 9.15 A code $F : J \times J' \rightarrow H$ is *quasi-permutable* if there exists a function $G : J \times J' \rightarrow J$ comonotonic with F such that

$$F(G(x, y), z) = F(G(x, z), y) \quad (x \in J; y, z \in J'). \quad (9.39)$$

We say in such a case that F is *permutable with respect to G* , or *G -permutable* for short. When $G = F$, that is, F is permutable with respect to itself, we simply say that F is *permutable*, a terminology consistent with Aczél (1966, chapter 6, p. 270).

Lemma 9.16 *A code $F : J \times J' \rightarrow H$ is G -permutable only if G is permutable, that is*

$$G(G(x, y), z) = G(G(x, z), y) \quad (x \in J; y, z \in J'). \quad (9.40)$$

The following results, due to Hosszú (1962a,b,c; cf. also Aczél, 1966, and Maksa, 2004, 2005), give the abstract representations for permutability and quasi-permutability.

Lemma 9.17 *Let J, J' and H be non-negative, real intervals of positive length.*

- (i) *A solvable code $F : J \times J' \rightarrow H$ is quasi-permutable if and only if there exist three continuous functions $m : \{f(x) + g(r) \mid x \in J, r \in J'\} \rightarrow H$, $f : J \rightarrow \mathbb{R}$, and $g : J' \rightarrow \mathbb{R}$, with m and f strictly increasing and g strictly monotonic, such that*

$$F(x, y) = m(f(x) + g(y)). \quad (9.41)$$

- (ii) *A code $G : J \times J' \rightarrow J$ is a permutable code if and only if, with f and g as above, we have*

$$G(x, y) = f^{-1}(f(x) + g(y)). \quad (9.42)$$

- (iii) *If a code $G : J \times J \rightarrow J$ is symmetric, then G is permutable if and only if there exists a strictly increasing and continuous function $f : J \rightarrow J$ satisfying*

$$G(x, y) = f^{-1}(f(x) + f(y)). \quad (9.43)$$

Many scientific laws embody permutable 2-codes, and so can be written in the form of Equations (9.42) or (9.43). It is easily checked that the Lorentz–FitzGerald contraction, for example, is permutable. We now give three more examples of permutable codes, and we give a counterexample that is not permutable.

Examples 9.18 (a) Beer's law. This law applies in a class of empirical situations where an incident radiation traverses some absorbing medium, so that only a fraction of the radiation goes through. In our notation, the expression of the law is

$$I(x, y) = xe^{-\frac{y}{c}} \quad (x, y \in \mathbb{R}_{++}, c \in \mathbb{R}_{++} \text{ constant}), \quad (9.44)$$

in which x denotes the intensity of the incident radiation, y is the concentration of the absorbing medium (which might be measured by the length of a tube containing the absorbing liquid), c is a reference level (such as the upper bound of the values of y), and $I(x, y)$ is the intensity of the transmitted radiation. Note that the function $I : \mathbb{R}_{++} \times \mathbb{R}_{++} \xrightarrow{\text{onto}} \mathbb{R}_{++}$ is a permutable code, as

$$I(I(x, y), z) = I(x, y)e^{-\frac{z}{c}} = xe^{-\frac{y}{c}}e^{-\frac{z}{c}} = I(I(x, z), y).$$

Our next two examples are permutable codes that appear in the context of Theorem 9.47, which examines a type of quasi-permutability.

(b) The volume of a cylinder. The permutability equation applies not only to many physical laws, but also to some fundamental formulas of geometry, such as the volume $C(\ell, r)$ of a cylinder of radius r and height ℓ . In this case, we have

$$C(\ell, r) = \ell\pi r^2, \quad (9.45)$$

which is easily shown to be permutable.

(c) The Pythagorean theorem. As another geometric example of a permutable code, consider the function

$$P(x, y) = \sqrt{x^2 + y^2} \quad (x, y \in \mathbb{R}_{++}), \quad (9.46)$$

representing the length of the hypotenuse of a right triangle in terms of the lengths of its sides. Indeed, this code is permutable, as

$$P(P(x, y), z) = \sqrt{P(x, y)^2 + z^2} = \sqrt{x^2 + y^2 + z^2} = P(P(x, z), y).$$

(d) The counterexample: van der Waals' equation.¹¹ One form of this equation is

$$T(p, v) = K \left(p + \frac{a}{v^2} \right) (v - b), \quad (9.47)$$

in which p is the pressure of a fluid, v is the volume of the container, T is the temperature, a and b are parameters, and K is the reciprocal of the Boltzmann constant. It is easily shown that the function T in (9.47) is not permutable.

9.8.3 The Translation Equation

Definition 9.19 A code $F : J \times J' \rightarrow J$ satisfies the *translation equation* (cf. Aczél, 1966, p. 245) or equivalently, is *translatable* if

$$F(F(x, y), z) = F(x, y + z) \quad (x \in J, y, z, y + z \in J'). \quad (9.48)$$

Note that Beer's law satisfies this property, as

$$I(I(x, y), z) = I(x, y)e^{-\frac{z}{c}} = xe^{-\frac{y}{c}}e^{-\frac{z}{c}} = xe^{-\frac{y+z}{c}} = I(x, y + z).$$

The reader may notice that Beer's law is a formalization of *exponential decay*. This observation gives insight into the translation equation. Consider the function given by $A(x, t) = xe^{-kt}$, in which x is an initial amount, t is elapsed time, $A(x, t)$ is the amount at time t , and k is a constant (called the *rate constant*). Note that this function satisfies the translation equation:

$$A(A(x, t), s) = A(x, t + s) \quad (x, t, s \in \mathbb{R}_+). \quad (9.49)$$

It makes sense that this exponential decay function would satisfy the translation equation: starting with an amount x and letting the decay process run for a time t , we get an amount $A(x, t)$. Using this as an initial amount and letting the process run for an additional time s , we get the same final amount as if we started with x and let the process run for a time of $t + s$.

We show in Theorem 9.49 that, under the meaningfulness condition on the corresponding collection of 2-codes, and assuming that a condition of *left homogeneity* is satisfied by at least one code, the exponential decay model of Beer's law is the only possible form for the collection of codes.

¹¹ Van der Waals' equation is a modification of the ideal gas law $PV = WRT$ in a situation where it is assumed that the molecules attract each other.

Here is the result giving the abstract representation for the translation equation (cf. Aczél, 1966, pp. 248–249).

Lemma 9.20 *Let $F : J \times J' \rightarrow H$ be a code such that $J' =]d, \infty[$ for some $d \in \mathbb{R}_+$, and suppose that, for some $a \in \mathbb{R}_+$, and for some $b \in \mathbb{R}_+$ or $b = \infty$, we have*

- (i) *either $J = [a, b[$ or $J =]a, b[$ if $F(x, y)$ is strictly increasing in y ;*
- (ii) *or $J =]a, b]$ or $J =]a, b[$ if $F(x, y)$ is strictly decreasing in y .*

Then, the code $F : J \times J' \rightarrow H$ is translatable if and only if there exists a function f satisfying the equation

$$F(x, y) = f(f^{-1}(x) + y).$$

9.8.4 The Bisymmetry Equation

Definition 9.21 Let J be an closed interval. A code $F : J \times J \rightarrow J$ is *bisymmetric*, or satisfies the *bisymmetry equation*, if for all x, y, z and w in J , we have

$$F(F(x, y), F(z, w)) = F(F(x, z), F(y, w)). \quad (9.50)$$

Lemma 9.22 (Aczél, 1966, pp. 278–287.) *Let J be a closed interval, and let $F : J \times J \rightarrow J$ be a code. Suppose moreover that $F(x, x) = x$ for all $x \in J$. Then the code F is bisymmetric if and only if, for some number q different from 0 or 1,*

$$F(x, y) = f((1 - q)f^{-1}(x) + qf^{-1}(y)) \quad (9.51)$$

for some continuous, strictly monotonic function f .

If moreover the code F is symmetric, then Equation (9.51) specializes into

$$F(x, y) = f\left(\frac{f^{-1}(x) + f^{-1}(y)}{2}\right). \quad (9.52)$$

The r.h.s. of Equation (9.52) is the so-called *quasi-arithmetic mean*.

9.8.5 The Autodistributivity Equations

Definition 9.23 A code $F : J \times J \rightarrow J$ is *autodistributive*, or satisfies the *autodistributivity equations*, if for any $x, y, z \in J$, the two equations below are satisfied:

$$\begin{aligned} F(F(x, y), z) &= F(F(x, z), F(y, z)) \\ F(x, F(y, z)) &= F(F(x, y), F(x, z)). \end{aligned} \quad (9.53)$$

In such a case, we can also say that *autodistributivity* holds for the code.

Interestingly, the abstract representation of autodistributive codes is the same as that of bisymmetric codes, as the next lemma states.

Lemma 9.24 (Aczél, 1966, p. 298.) *Let a code $F : J \times J \rightarrow J$ be reducible on both sides. Then the code F is autodistributive if and only if, for some number q different from 0 or 1, we have*

$$F(x, y) = f((1 - q)f^{-1}(x) + qf^{-1}(y)) \quad (9.54)$$

for some strictly monotonic function f .

If moreover the code F is symmetric, then Equation (9.54) specializes into

$$F(x, y) = f\left(\frac{f^{-1}(x) + f^{-1}(y)}{2}\right). \quad (9.55)$$

9.8.6 The Abstract Lorentz–FitzGerald–Doppler Equations

Definition 9.25 Let $L : \mathbb{R}_{++} \times [0, c] \rightarrow \mathbb{R}_{++}$ be a code, with c a positive real number, and let $\oplus : [0, c] \times [0, c] \rightarrow [0, c]$ be an operation. The pair (L, \oplus) is an *abstract LFD-pair* if it satisfies the following five conditions.¹²

1. The function L is strictly increasing in the first variable, strictly decreasing in the second variable, continuous in both variables, and for all $\lambda, \lambda' \in \mathbb{R}_+$ and $v, v' \in [0, c]$, and for any $\alpha > 0$, we have

$$L(\lambda, v) \leq L(\lambda', v') \iff L(\alpha\lambda, v) \leq L(\alpha\lambda', v').$$

2. $L(\lambda, 0) = \lambda$ for all $\lambda \in \mathbb{R}_+$.
3. $\lim_{v \rightarrow c} L(\lambda, v) = 0$.
4. The operation \oplus is continuous, strictly increasing in both variables, and has 0 as an identity element.
5. Either Axiom [R] or Axiom [M] below is satisfied:

$$[\text{R}] \quad L(L(\lambda, v), w) = L(\lambda, v \oplus w) \quad (\lambda > 0, \text{ and } v, w \in [0, c]).$$

$$[\text{M}] \quad L(\lambda, v) \leq L(\lambda', v') \iff L(\lambda, v \oplus w) \leq L(\lambda', v' \oplus w) \\ (\lambda, \lambda' > 0, \text{ and } v, v', w \in [0, c]).$$

In words, in the context of the Lorentz–FitzGerald–Doppler equations in special relativity, Axioms [R] and [M] state the following ideas.

Axiom [R]: *One iteration of the function L involving two velocities v and w has the same effect on the perceived length as adding v and w via the operation \oplus .*

Axiom [M]: *Adding a velocity via the operation \oplus preserves the order of the function L .*

The following representation theorem for abstract LFD-pairs is due to Falmagne and Doignon (2010).

¹² We use λ rather than ℓ for the first variable in this definition because we jointly deal with both the Doppler effect equation (9.56) and the Lorentz–FitzGerald equation (9.32).

Lemma 9.26 *If (L, \oplus) is an abstract LFD-pair, then the following equivalences hold:*

$$[\text{R}] \iff ([\text{DE}^\dagger] \& [\text{AV}^\dagger]) \iff [\text{M}],$$

with for some strictly increasing function u and some positive constant ξ :

$$\begin{aligned} [\text{DE}^\dagger] \quad L(\lambda, v) &= \lambda \left(\frac{c-u(v)}{c+u(v)} \right)^{\frac{\xi}{2}}; \\ [\text{AV}^\dagger] \quad v \oplus w &= u^{-1} \left(\frac{u(v)+u(w)}{1+\frac{u(v)u(w)}{c^2}} \right). \end{aligned}$$

Note that with $\xi = \frac{1}{2}$ and if u is the identity function, $[\text{DE}^\dagger]$ becomes the Doppler effect equation

$$[\text{DE}] \quad L(\lambda, v) = \lambda \left(\frac{c-v}{c+v} \right)^{\frac{1}{2}}. \quad (9.56)$$

9.9 Defining Ratio Scale Meaningfulness

9.9.1 Motivation

In this section, we turn to defining a particular type of invariance, a meaningfulness condition, whose consequence is that the form of an expression representing a scientific or geometric law would not be altered with a change in measurement units. The definition (Definition 9.9.2), which applies to n -codes regarded as functions of n real, ratio scale variables, follows Falmagne and Doble (2015) and generalizes that used by Falmagne (2004).¹³

In Subsection 9.7.1, we gave an example of such a condition. The example dealt with the side lengths of right triangles, a context in which all of the measurement variables were identical ratio scales. Here we give further motivation through a second example, one in which not all variables are measured with the same scale.

Consider the Lorentz–FitzGerald contraction mentioned earlier.¹⁴ We recall that this term denotes a phenomenon in special relativity, according to which the apparent length of a rod measured by an observer moving on a line parallel to the rod, at the speed v with respect to that rod, is a decreasing function of v , vanishing as v approaches the speed of light. This function is

$$L(\ell, v) = \ell \sqrt{1 - \left(\frac{v}{c} \right)^2}, \quad (9.57)$$

in which $c > 0$ denotes the speed of light, ℓ is the actual length of the rod (for an observer at rest with respect to the rod), and $L : \mathbb{R}_+ \times [0, c[\xrightarrow{\text{onto}} \mathbb{R}_+$ is the length of the rod measured by the moving observer.

¹³ See also Falmagne and Narens (1983), Narens (2002a, 2007), and Falmagne (2015).

¹⁴ In the rest of this subsection, our discussion follows closely Falmagne (2004).

Although standard for a mathematical or scientific context, this notation is ambiguous, as the units of ℓ and of v are not specified. For example, writing $L(70, 3)$ has no empirical meaning; it is necessary to specify the units, e.g., 70 “meters” and 3 “kilometers per second.”

Recall that our goal is to make precise an invariance with respect to any change in the units. So, as in the right triangle example in Subsection 9.7.1, we again indicate the measurement unit explicitly in the notation. (See also Suppes, 2002, p. 120, in this regard.) In particular, we consider $L(\ell, v)$ to be shorthand for $L_{1,1}(\ell, v)$, in which the $_{1,1}$ indices of $L_{1,1}$ signify the *initial* or *anchor* units, which are fixed arbitrarily. These initial units could be, for example, m (meter) and km/s . A change of units amounts to multiplying ℓ and v in each pair (ℓ, v) by some positive constants α and β , respectively, and also multiplying L by α and the speed of light c by β . Doing so defines a new function $L_{\alpha,\beta}$, which is different from $L = L_{1,1}$ if either $\alpha \neq 1$ or $\beta \neq 1$ (or both), but carries the same empirical information. For example, if our new units are km and m/s , then the two expressions

$$L_{10^{-3}, 10^3}(.07, 3000) \quad \text{and} \quad L(70, 3) = L_{1,1}(70, 3),$$

while numerically not equal, describe the same empirical situation.

With this in mind, we would have for $L_{\alpha,\beta}$ that

$$L_{\alpha,\beta}(\ell, v) = \ell \sqrt{1 - \left(\frac{v}{\beta c}\right)^2}, \tag{9.58}$$

and L and $L_{\alpha,\beta}$ would be related by

$$\frac{1}{\alpha} L_{\alpha,\beta}(\alpha \ell, \beta v) = \left(\frac{1}{\alpha}\right) \alpha \ell \sqrt{1 - \left(\frac{\beta v}{\beta c}\right)^2} = \ell \sqrt{1 - \left(\frac{v}{c}\right)^2} = L(\ell, v).$$

Then, for any α, β, v and μ in \mathbb{R}_{++} , we would have

$$\frac{1}{\alpha} L_{\alpha,\beta}(\alpha \ell, \beta v) = \frac{1}{\nu} L_{\nu,\mu}(\nu \ell, \mu v). \tag{9.59}$$

Equation (9.59) gives an invariance condition that is generalized in Definition 9.27 below as a definition of meaningfulness; see Equation (9.61).

As previewed earlier, and as discussed in this example and in the example in Subsection 9.7.1, the definition of meaningfulness described in Part II applies to a collection of codes, rather than to a single code. In the current example, the collection is $\{L_{\alpha,\beta} \mid (\alpha, \beta) \in \mathbb{R}_{++}^2\}$, in which each (α, β) corresponds to a different choice of units.

We now define the concept of meaningfulness in the general case of a family of n -codes.

9.9.2 Defining Meaningfulness

Definition 9.27 Let J_1, \dots, J_{n+1} be $n+1$ non-negative, real intervals, and suppose that

$$\mathcal{F} = \{F_\alpha \mid \alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}_{++}^n\} \quad (9.60)$$

is a collection of n -codes, with for the *initial code* F

$$F = F_{\underbrace{1, \dots, 1}_{n \text{ indices}}} : J_1 \times \dots \times J_n \xrightarrow{\text{onto}} J_{n+1}.$$

Each of the terms $\alpha_1, \dots, \alpha_n$ in the vector α indexing a code F_α in \mathcal{F} represents a change of the unit of one of the measurement scales. We will specify the domain and range of any code F_α in a moment.

Let $\delta_1, \dots, \delta_n$ be a finite sequence of rational numbers. The collection of n -codes \mathcal{F} is $(\delta_1, \dots, \delta_n)$ -meaningful if for any vector

$$(x_1, \dots, x_n) \in J_1 \times \dots \times J_n$$

and any vectors $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}_{++}^n$ and $\mu = (\mu_1, \dots, \mu_n) \in \mathbb{R}_{++}^n$, the following equality holds:

$$\frac{1}{\prod_{i=1}^n \alpha_i^{\delta_i}} F_\alpha(\alpha_1 x_1, \dots, \alpha_n x_n) = \frac{1}{\prod_{i=1}^n \mu_i^{\delta_i}} F_\mu(\mu_1 x_1, \dots, \mu_n x_n), \quad (9.61)$$

which implies

$$\frac{1}{\prod_{i=1}^n \alpha_i^{\delta_i}} F_\alpha(\alpha_1 x_1, \dots, \alpha_n x_n) = F_{\underbrace{1, \dots, 1}_{n \text{ indices}}}(x_1, \dots, x_n) \quad (9.62)$$

$$= F(x_1, \dots, x_n), \quad (9.63)$$

yielding

$$F_\alpha(\alpha_1 x_1, \dots, \alpha_n x_n) = \prod_{i=1}^n \alpha_i^{\delta_i} F(x_1, \dots, x_n). \quad (9.64)$$

So, for example, if the intervals J_1, \dots, J_{n+1} are, respectively, defined by $[a_1, a'_1], \dots, [a_{n+1}, a'_{n+1}]$, then any code F_α in a $(\delta_1, \dots, \delta_n)$ -meaningful family \mathcal{F} is a mapping

$$F_\alpha : [\alpha_1 a_1, \alpha_1 a'_1] \times \dots \times [\alpha_n a_n, \alpha_n a'_n] \longrightarrow \left[\left(\prod_{i=1}^n \alpha_i^{\delta_i} \right) a_{n+1}, \left(\prod_{i=1}^n \alpha_i^{\delta_i} \right) a'_{n+1} \right].$$

This definition can be adapted in an obvious way when some or all of the ratio scales measuring the variables are the same. For an example, see the case of two identical variables in Equation (9.67).

Here is the concept of self-transforming (see earlier) in the context of this meaningfulness condition.

Definition 9.28 A meaningful collection of n -codes \mathcal{F} is *self-transforming* if for any code F_{α} in the collection, the measurement unit of the output of the code F_{α} – the value of the function F_{α} – is the same as the measurement unit of its first variable. For example: for every vector $\alpha = (\alpha_1, \dots, \alpha_n)$ in \mathbb{R}_{++}^n , we have

$$\delta_1 = 1 \text{ and } \delta_2 = \delta_3 = \dots = \delta_n = 0,$$

and so

$$\alpha_1^{\delta_1} \times \dots \times \alpha_n^{\delta_n} = \alpha_1.$$

In such a case, we may say that \mathcal{F} is an *ST-collection*. For convenience, we will sometimes say that \mathcal{F} is *ST-meaningful*.

If the ST-collection \mathcal{F} is $(1, 0, \dots, 0)$ -meaningful, then Equations (9.61)–(9.63) simplify into

$$\frac{1}{\alpha_1} F_{\alpha}(\alpha_1 x_1, \dots, \alpha_n x_n) = \frac{1}{\mu_1} F_{\mu}(\mu_1 x_1, \dots, \mu_n x_n) = F(x_1, \dots, x_n).$$

This was the case with the Lorentz–FitzGerald contraction, Equation (9.59), which had $n = 2$.

In what follows in Part II, we mostly apply these concepts to ST-collections of 2-codes $\mathcal{F} = \{F_{\alpha, \beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$. (Exceptions appear as the monomial laws, seen in Example 9.29(a) and Theorem 9.86.) For a 2-code, we consider three real, non-negative intervals $[a, a'[$, $[b, b'[$ and $[d, d'[$.

For the initial code F , we have

$$F = F_{1,1} : [a, a'[\times[b, b'[\xrightarrow{\text{onto}} [d, d'[,$$

and so, for any code $F_{\alpha, \beta}$ in \mathcal{F} , we have

$$F_{\alpha, \beta} : [\alpha a, \alpha a'[\times[\beta b, \beta b'[\xrightarrow{\text{onto}} [\alpha^{\delta_1} \beta^{\delta_2} d, \alpha^{\delta_1} \beta^{\delta_2} d'[, \quad (\alpha, \beta \in \mathbb{R}_{++}).$$

The meaningfulness equation (9.61) specializes into

$$\frac{1}{\alpha^{\delta_1} \beta^{\delta_2}} F_{\alpha, \beta}(\alpha x, \beta r) = \frac{1}{\mu^{\delta_1} \nu^{\delta_2}} F_{\mu, \nu}(\mu x, \nu r), \quad (x \in [a, a'[; r \in [b, b'[), \quad (9.65)$$

which implies

$$F_{\alpha, \beta}(\alpha x, \beta r) = \alpha^{\delta_1} \beta^{\delta_2} F(x, r). \quad (9.66)$$

When the same measurement scale is used for the two variables in a 2-code, we write $F_{\alpha,\alpha} = F_\alpha$. Equation (9.65) becomes then

$$\frac{1}{\alpha^{\delta_1+\delta_2}} F_\alpha(\alpha x, \alpha r) = \frac{1}{\mu^{\delta_1+\delta_2}} F_\mu(\mu x, \mu r), \quad (x \in [a, a'[, r \in [a, a']). \quad (9.67)$$

We now examine these definitions through examples, several of which are used throughout Part II.

9.9.3 Applications

Examples 9.29 (a) A product of powers. Here, the collection $\mathcal{F} = \{F_\alpha\}$ is $(\delta_1, \dots, \delta_n)$ -meaningful, with $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}_{++}^n$. Each code F_α is a product of powers, with the initial code $F(x_1, \dots, x_n) = \prod_{i=1}^n x_i^{\delta_i}$. For any $\mu = (\mu_1, \dots, \mu_n)$ in \mathbb{R}_{++}^n , we have

$$\begin{aligned} \frac{1}{\prod_{i=1}^n \mu_i^{\delta_i}} F_\mu(\mu_1 x_1, \dots, \mu_n x_n) &= \frac{1}{\prod_{i=1}^n \mu_i^{\delta_i}} \prod_{i=1}^n (\mu_i x_i)^{\delta_i} \\ &= \prod_{i=1}^n x_i^{\delta_i} = F(x_1, \dots, x_n) \\ &= \frac{1}{\prod_{i=1}^n \alpha_i^{\delta_i}} F_\alpha(\alpha_1 x_1, \dots, \alpha_n x_n). \end{aligned}$$

(b) The Lorentz–FitzGerald contraction. The collection of codes is $\mathcal{L} = \{L_{\alpha,\beta} \mid (\alpha, \beta) \in \mathbb{R}_{++}^2\}$. Requiring that the collection \mathcal{L} be ST-meaningful, we have

$$\frac{1}{\alpha^1 \beta^0} L_{\alpha,\beta}(\alpha \ell, \beta v) = \left(\frac{1}{\alpha} \right) \alpha \ell \sqrt{1 - \left(\frac{\beta v}{\beta c} \right)^2} = \ell \sqrt{1 - \left(\frac{v}{c} \right)^2} = L(\ell, v),$$

and so

$$\frac{1}{\alpha} L_{\alpha,\beta}(\alpha \ell, \beta v) = \frac{1}{v} L_{v,\mu}(v \ell, \mu v).$$

Note that this is Equation (9.59), which is a special case of (9.61) and (9.65). The family \mathcal{L} is self-transforming. So is the family of our next example.

(c) Beer's law. We encountered this law as Example 9.18(a). We have a collection $\mathcal{I} = \{I_{\alpha,\beta} \mid (\alpha, \beta) \in \mathbb{R}_{++}^2\}$ of codes. Assuming the collection to be ST-meaningful gives

$$\frac{1}{\alpha^1 \beta^0} I_{\alpha,\beta}(\alpha x, \beta y) = \frac{1}{\alpha} (\alpha x) e^{-\frac{\beta y}{\beta c}} = x e^{-\frac{y}{c}} = I(x, y),$$

yielding

$$\frac{1}{\alpha} I(\alpha x, \beta y) = \frac{1}{v} I(v x, \mu y),$$

another special case of (9.61) and (9.65).

(d) The volume of a cylinder. This example involves a collection of codes $\mathcal{C} = \{C_\alpha \mid \alpha \in \mathbb{R}_{++}\}$, which must be $(1, 2)$ -meaningful. We get

$$\frac{1}{\alpha^1 \alpha^2} C_\alpha(\alpha \ell, \alpha r) = \frac{1}{\alpha^3} (\alpha \ell) \pi (\alpha r)^2 = \ell \pi r^2 = C(\ell, r),$$

yielding

$$\frac{1}{\alpha^3} C_\alpha(\alpha \ell, \alpha r) = \frac{1}{v^3} C_v(v \ell, v r).$$

The collection \mathcal{C} is not an ST-collection since $\alpha^3 \neq \alpha$.

(e) The Pythagorean theorem. Here, we have only one measurement scale, which is the same for the two input variables and for the output variable. We require the collection of codes $\mathcal{P} = \{P_\alpha \mid \alpha \in \mathbb{R}_{++}\}$ to be $(\frac{1}{2}, \frac{1}{2})$ -meaningful (with $P_\alpha = P_{\alpha, \alpha}$). We obtain

$$\frac{1}{\alpha^{\frac{1}{2}} \alpha^{\frac{1}{2}}} P_\alpha(\alpha x, \alpha y) = \frac{1}{\alpha} P_\alpha(\alpha x, \alpha y) \quad (9.68)$$

$$= \frac{1}{\alpha} \sqrt{(\alpha x)^2 + (\alpha y)^2} = \sqrt{x^2 + y^2} = P(x, y). \quad (9.69)$$

9.9.4 The Falmagne and Narens Definition

The definition of meaningfulness given above in Definition 9.27 originated with a definition by Falmagne and Narens (1983). This latter definition applies to both ratio scales and other types of scales, and it is weaker than Definition 9.27 (see Theorem 9.35). The two definitions have in common that they apply to a collection of functions, each of which is indexed by the units of the measurement scales, rather than to a single function.

There are two related definitions of meaningfulness in the sense of Falmagne and Narens (1983): “one-to-one meaningfulness” and “order-meaningfulness.”

Here is an example illustrating “order-meaningfulness.” (See also Doble, 2002.)

Example 9.30 The attractive (gravitational) force F between two objects, one of mass m_1 and the other of mass m_2 , at a distance r between their centers of mass, is given by the equation

$$F(m_1, m_2, r) = G m_1 m_2 \frac{1}{r^2}, \quad (9.70)$$

in which G is a “dimensional constant” (the universal gravitational constant). Note that the numerical value of G depends on the units used in the measurement of the variables. Suppose there is some choice of an initial unit for m_1 and m_2 , and some choice of an initial unit for r , and that we change this pair of units via respective multiplications by α and β . If we define

$$f : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++} : x \mapsto \alpha x, \quad g : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++} : x \mapsto \beta x,$$

then Equation (9.70) becomes

$$F_{f,f,g}(m_1, m_2, r) = G(f, f, g)m_1 m_2 \frac{1}{r^2}, \quad (9.71)$$

which indicates the dependence on the units being used.¹⁵ Suppose that we measure the attractive force for two different triples of inputs (m_1, m_2, r) and (m'_1, m'_2, r') , using the scales f and g , and that the first attractive force does not exceed the second. The Falmagne and Narens (1983) definition of “order-meaningfulness” says that this relationship between the forces should hold even if we use different scales f^* and g^* , that is,

$$\begin{aligned} F_{f,f,g}(f(m_1), f(m_2), g(r)) &\leq F_{f,f,g}(f(m'_1), f(m'_2), g(r')) \\ \iff & \\ F_{f^*,f^*,g^*}(f^*(m_1), f^*(m_2), g^*(r)) &\leq F_{f^*,f^*,g^*}(f^*(m'_1), f^*(m'_2), g^*(r')). \end{aligned} \quad (9.72)$$

Note that the formula in Equation (9.71) satisfies this requirement:

$$\begin{aligned} F_{f,f,g}(f(m_1), f(m_2), g(r)) &\leq F_{f,f,g}(f(m'_1), f(m'_2), g(r')) \\ \iff & \\ G(f, f, g)\alpha m_1 \alpha m_2 \frac{1}{(\beta r)^2} &\leq G(f, f, g)\alpha m'_1 \alpha m'_2 \frac{1}{(\beta r')^2} \\ \iff & \\ m_1 m_2 \frac{1}{r^2} &\leq m'_1 m'_2 \frac{1}{r'^2}, \end{aligned}$$

and this last inequality does not depend on the units of the scales used. So, (9.72) is satisfied for any functions f , g , f^* and g^* specifying the units. This shows that the collection of codes given by (9.71) is “order-meaningful” in the sense of Falmagne and Narens (1983).

Here are the definitions used by Falmagne and Narens (1983).

Definition 9.31 Let $\mathcal{F} = \{f \mid f : J \rightarrow \mathbb{R}_{++}\}$ and $\mathcal{G} = \{g \mid g : J' \rightarrow \mathbb{R}_{++}\}$ be two families of *scales* with domains J and J' , respectively, with both f and g strictly increasing and continuous for all $f \in \mathcal{F}$ and $g \in \mathcal{G}$. Let $R \subseteq \mathcal{F} \times \mathcal{G}$ be some binary relation such that $(\iota_J, \iota_{J'}) \in R$, where ι_J and $\iota_{J'}$ are the identity functions on J and J' , respectively. With $(f, g) \in R$, suppose that $M_{f,g} : f(J) \times g(J') \rightarrow \mathbb{R}$ is continuous, strictly increasing in its first argument, and strictly monotonic in its second argument. Then the family $\mathcal{M} = \{M_{f,g} \mid (f, g) \in R\}$ is a *family of (numerical) codes*. Each $M_{f,g} \in \mathcal{M}$ is a *(numerical) code*.

Definition 9.32 Let $\mathcal{M} = \{M_{f,g} \mid (f, g) \in R\}$ be a family of codes, with $f : J \rightarrow \mathbb{R}_{++}$ and $g : J' \rightarrow \mathbb{R}_{++}$ strictly increasing and continuous for any

¹⁵ Note that the writing conventions of Equation (9.71) are slightly different from those of our meaningfulness equations in Definition 9.27, in that the latter would lead us to write $F_{\alpha,\alpha,\beta}(m_1, m_2, r) = G(\alpha, \alpha, \beta)m_1 m_2 \frac{1}{r^2}$.

$(f, g) \in R$. The family \mathcal{M} is *one-to-one meaningful* (or *1-1-meaningful* for short) if whenever $(f, g) \in R$ and $(f^*, g^*) \in R$, we have

$$\begin{aligned} M_{f,g}(f(a), g(x)) &= M_{f,g}(f(b), g(y)) \\ \iff \\ M_{f^*,g^*}(f^*(a), g^*(x)) &= M_{f^*,g^*}(f^*(b), g^*(y)) \end{aligned} \tag{9.73}$$

for all $a, b \in J$ and $x, y \in J'$. The family \mathcal{M} is *order-meaningful* if whenever $(f, g) \in R$ and $(f^*, g^*) \in R$, we have, for all $a, b \in J$ and $x, y \in J'$,

$$\begin{aligned} M_{f,g}(f(a), g(x)) &\leq M_{f,g}(f(b), g(y)) \\ \iff \\ M_{f^*,g^*}(f^*(a), g^*(x)) &\leq M_{f^*,g^*}(f^*(b), g^*(y)). \end{aligned} \tag{9.74}$$

The following theorem gives a useful criterion for 1-1-meaningfulness or order-meaningfulness. See Falmagne and Narens (1983, p. 298) for a sketch of the proof.

Theorem 9.33 *A family of codes $\mathcal{M} = \{M_{f,g} \mid (f, g) \in R\}$ is 1-1-meaningful (respectively, order-meaningful) if and only if, for all $(f, g) \in R$, there exists a one-to-one (respectively strictly increasing, continuous) function $H_{f,g}$ mapping the range of $M = M_{i,i}$ onto the range of $M_{f,g}$ such that*

$$H_{f,g}[M(a, x)] = M_{f,g}[f(a), g(x)] \tag{9.75}$$

whenever both members are defined.

Theorem 9.35 below says that, for ratio scales, meaningfulness in the sense of Definition 9.27 implies both 1-1-meaningfulness and order-meaningfulness. Here is an example (from Falmagne and Doble, 2015, pp. 68–69) that shows the converse does not hold.

Example 9.34 Consider the family $\mathcal{M} = \{M_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$, a ratio scale family of 2-codes, in which

$$M_{\alpha,\beta} : \mathbb{R}_{++} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$$

is defined by

$$M_{\alpha,\beta}(a, x) = e^{\alpha x}.$$

Writing $M_{1,1} = M$, we have that $M = M_{\alpha,\beta}$ for all $\alpha, \beta > 0$. Also note that

$$M_{\alpha,\beta}(\alpha a, \beta x) = e^{\alpha a \beta x} = [M(a, x)]^{\alpha \beta}. \tag{9.76}$$

By Theorem 9.33, \mathcal{M} is 1-1-meaningful (respectively, order-meaningful) since there exists a one-to-one (respectively, a strictly increasing, continuous) function $H_{\alpha,\beta}$ mapping the range of M onto the range of $M_{\alpha,\beta}$ such that

$$H_{\alpha,\beta}(M(a, x)) = M_{\alpha,\beta}(\alpha a, \beta x). \tag{9.77}$$

From (9.76), we see that $H_{\alpha,\beta}(s) = s^{\alpha \beta}$ for all $s \in]1, \infty[$.

However, \mathcal{M} is not (δ_1, δ_2) -meaningful for any $\delta_1, \delta_2 \in \mathbb{Q}$, as we now show. If it were, then we would have for all $a, x, \alpha, \beta \in \mathbb{R}_{++}$ and some $\delta_1, \delta_2 \in \mathbb{Q}$:

$$\alpha^{\delta_1} \beta^{\delta_2} M(a, x) = M_{\alpha, \beta}(\alpha a, \beta x), \quad (9.78)$$

that is,

$$\alpha^{\delta_1} \beta^{\delta_2} e^{ax} = (e^{ax})^{\alpha \beta}. \quad (9.79)$$

But (9.79) cannot hold for all $a, x, \alpha, \beta \in \mathbb{R}_{++}$.

Indeed, dividing both sides by $e^{ax} > 0$, we get

$$\alpha^{\delta_1} \beta^{\delta_2} = e^{ax(\alpha \beta - 1)}. \quad (9.80)$$

For $\alpha \beta \neq 1$, the r.h.s. of (9.80) varies with a and x , while the l.h.s. does not.

Theorem 9.35 *Let $\mathcal{M} = \{M_{\alpha, \beta}\}$ be a ratio scale family of 2-codes, with $M : J \times J' \rightarrow \mathbb{R}$ for real, positive, possibly infinite intervals J and J' . If the family \mathcal{M} is (δ_1, δ_2) -meaningful for some $\delta_1, \delta_2 \in \mathbb{Q}$ in the sense of Definition 9.27, then it is also order-meaningful (respectively, 1-1-meaningful). (Example 9.34 shows that the converse does not hold.)*

For a proof, see Falmagne and Doble (2015, p. 69). Note that meaningfulness in the sense of Definition 9.27 follows from the special case of 1-1-meaningfulness in which the function $H_{\alpha, \beta}$ of Theorem 9.33 is defined by $H_{\alpha, \beta}(s) = \alpha^{\delta_1} \beta^{\delta_2} s$.

Falmagne and Narens (1983) also investigated two other forms of invariance different from, but closely related to, order-meaningfulness. We define them below.

Definition 9.36 Suppose \mathcal{F} and \mathcal{G} are scale families with domains J and J' , respectively, and take $R \subseteq \mathcal{F} \times \mathcal{G}$. A family $\mathcal{M} = \{M_{f, g} \mid (f, g) \in R\}$ of codes is *dimensionally order-invariant*¹⁶ if, whenever $(f^*, g^*) \in R$, we have, for all $a, b \in J$ and $x, y \in J'$,

$$\begin{aligned} M_{f, g}(f^*(a), g^*(x)) \leq M_{f, g}(f^*(b), g^*(y)) \\ \iff \\ M_{f, g}(a, x) \leq M_{f, g}(b, y). \end{aligned} \quad (9.81)$$

Definition 9.37 A family $\mathcal{M} = \{M_{f, g} \mid (f, g) \in R\}$ of codes is *isotone* if, for any $(f, g) \in R$, there exists a strictly increasing, continuous function $m_{f, g}$ such that, for all $a, b \in J$ and $x, y \in J'$, we have

$$M_{f, g}(a, x) = m_{f, g}[M(a, x)] \quad (9.82)$$

Falmagne and Narens (1983) showed that the three conditions of isotonicity (9.82), order-meaningfulness (9.74), and dimensional order-invariance (9.81) are independent, but that any two of them imply the third (see also Luce *et al.*, 1990, p. 315, theorem 6).

The goal of Falmagne and Narens (1983) was the same as that pursued by many of the researchers mentioned in Section 9.7, that is: to derive, *a priori*, the possible functional forms of scientific laws. The following are two results they obtained.

¹⁶ The term used by Falmagne and Narens (1983) is “*dimensionally invariant*.”

Theorem 9.38 (Falmagne & Narens, 1983, corollary 5, p. 308.)¹⁷ *Let \mathcal{F} be a ratio scale family on J , and let \mathcal{G} be a ratio scale family on J' . Suppose that $\mathcal{M} = \{M_{f,g} \mid f \in \mathcal{F}, g \in \mathcal{G}\}$ is an isotone, order-meaningful family. Then, there exist constants $\theta > 0$ and $\delta \neq 0$ such that, for all $a \in J$ and $x \in J'$, and with a strictly increasing, continuous function G , we have*

$$M(a, x) = G(a^\theta x^\delta).$$

Theorem 9.39 (Falmagne & Narens, 1983, theorem 8, p. 311.) *Let \mathcal{F} be a ratio scale family, and let $\mathcal{M} = \{M_{f,f} \mid f \in \mathcal{F}\}$ be an isotone, order-meaningful family such that $M(a, x) = F(u(a)h(x))$ for continuous u , h , and F , with F strictly increasing. Then either*

$$M(a, b) = F(a^\theta b^\delta) \quad \text{or} \quad M(a, b) = F(\tau a^\theta + \xi b^\theta),$$

with $\theta, \tau > 0$ and $\delta, \xi \neq 0$ constants.

These results are in the same vein as the ones in Section 9.11 below, but weaker, in that they involve unknown functions F and G . By comparison, see, for example, Theorems 9.45 and 9.47.

9.10 Propagating Abstract Axioms via Meaningfulness

We again make the observation that the approach to measurement invariance based on Definition 9.27 of this chapter, like the approach of Falmagne and Narens (1983) described above, places families of scientific functions as the key objects of study, rather than single functions.¹⁸ The mathematical results obtained are powerful, at least inasmuch as they pertain to large collections. A potential obstacle could be that the assumptions entering into the results would have to be strong, as these assumptions would have to apply to entire families. However, it turns out that the condition in Definition 9.27 has the remarkable feature that, in several important cases, a property imposed on a single code of a meaningful collection may be automatically propagated to all the codes in the collection.

What follows are three lemmas illustrating this feature of the meaningfulness condition.¹⁹ These lemmas show that a property of a single code in a (δ_1, δ_2) -meaningful collection of 2-codes may automatically extend to all the codes in the collection. For the proofs of Lemmas 9.40 and 9.41, see Falmagne and Doble (2015, pp. 80–82). (The proof of Lemma 9.44 is included here.)

17 See also Luce *et al.* (1990, theorem 7, p. 315).

18 For further discussion of the role of families of functions in the study of invariance, see Narens (2002a, pp. 64–65 and 254–256).

19 These lemmas are due to Falmagne (2015).

9.10.1 Propagating Solvability and other Conditions

Lemma 9.40 Let $\mathcal{F} = \{F_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ be a (δ_1, δ_2) -meaningful collection of codes. Suppose that some code $F_{\alpha,\beta}$ in the collection \mathcal{F} satisfies any of the following four properties:

- (i) $F_{\alpha,\beta}$ is solvable (Definition 9.11);
- (ii) $F_{\alpha,\beta}$ is differentiable in both variables;
- (iii) $F_{\alpha,\beta}$ is quasi-permutable (Definition 9.15);
- (iv) with $\alpha = \beta$ (that is, $F_{\alpha,\beta} = F_{\alpha,\alpha} = F_\alpha$), $F_{\alpha,\beta}$ is a symmetric function.

Then all the codes in \mathcal{F} satisfy the same property.

9.10.2 The Self-transforming Case

Lemma 9.41 Let $\mathcal{F} = \{F_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ be a ST-collection of codes which is (δ_1, δ_2) -meaningful. Suppose that some code $F_{\alpha,\beta}$ in the collection \mathcal{F} satisfies any of the following four properties:

- (i) $F_{\alpha,\beta}$ is translatable;
- (ii) $F_{\alpha,\beta}$ is symmetric, assuming $\alpha = \beta$ (so $F_{\alpha,\beta} = F_\alpha$), and is associative;
- (iii) $F_{\alpha,\beta}$ is permutable.
- (iv) and for collections such that $F_\alpha = F_{\alpha,\beta}$, the function F_α is homogeneous, that is, $\gamma F_\alpha(x, r) = F_\alpha(\gamma x, \gamma r)$ for any $\gamma > 0$.

Then all the codes in \mathcal{F} satisfy the same property. Moreover, Condition (iv) implies that $F_\alpha(x, r) = F_1(x, r) = F(x, r)$ for all $\alpha > 0$ and all (x, r) in the domain of F .

9.10.3 The Meaningful Lorentz–FitzGerald–Doppler Systems

Definition 9.42 Let $\mathcal{L} = \{L_{\mu,v} \mid \mu, v \in \mathbb{R}_{++}\}$ be a ST-meaningful collection of codes, with $L_{\mu,v} : \mathbb{R}_{++} \times [0, c[\xrightarrow{\text{onto}} \mathbb{R}_{++}$ and $c \in \mathbb{R}_{++}$. Let $\mathcal{O} = \{\oplus_v \mid v \in \mathbb{R}_{++}\}$ be a $(1, 0)$ -meaningful collection of operators, with

$$\oplus_v : [0, c[\times [0, c[\xrightarrow{\text{onto}} [0, c[\quad \text{and} \quad v \oplus_v w = v \left(\frac{v}{v} \oplus \frac{w}{v} \right)$$

(with $v \in \mathbb{R}_{++}$, $v, w \in [0, c[$).

Suppose that each code $L_{\mu,v} \in \mathcal{L}$ is paired with a binary operation $\oplus_v \in \mathcal{O}$, forming an ordered pair $(L_{\mu,v}, \oplus_v)$, with the initial ordered pair $(L_{1,1}, \oplus_1) = (L, \oplus)$. Then the pair of collections $(\mathcal{L}, \mathcal{O})$ is called a *meaningful LFD system*.

Remark 9.43 In the proof of the next lemma, the first equation is

$$L_{\alpha,\beta}(\lambda, v) = \alpha L \left(\frac{\lambda}{\alpha}, \frac{v}{\beta} \right) \tag{9.83}$$

which is equivalent to

$$L_{\alpha,\beta}(\alpha\lambda, \beta v) = \alpha L(\lambda, v). \tag{9.84}$$

In Equation (9.84), we have $L : \mathbb{R}_+ \times [0, c[\rightarrow \mathbb{R}_+$. But in the r.h.s. of Equation (9.83), we cannot have $\frac{v}{\beta} \in [0, c[$. The reason is that $0 < \beta$ implies

$$0 \leq v < c \iff 0 \leq \frac{v}{\beta} < \frac{c}{\beta}.$$

So, assuming that $\frac{v}{\beta} \in [0, c[$ would lead to a contradiction: with $1 < \beta$, we could get $c < v < \beta c$. So, the upper bound of the second variable in $L(\frac{\lambda}{\alpha}, \frac{v}{\beta})$ is now $\frac{c}{\beta}$.

Propagation lemma for abstract LFD-pairs 9.44 *Suppose that some ordered pair $(L_{\mu,\nu}, \oplus_\nu)$ from a meaningful LFD system $(\mathcal{L}, \mathcal{O})$ is an abstract LFD-pair, that is, $(L_{\mu,\nu}, \oplus_\nu)$ satisfies Conditions 1–5 of the definition of an abstract LFD-pair. Then any ordered pair $(L_{\alpha,\beta}, \oplus_\beta)$, with $L_{\alpha,\beta} \in \mathcal{L}$ and $\oplus_\beta \in \mathcal{O}$, is also such an abstract LFD-pair.*

So, meaningfulness enables the propagation of all five conditions to any ordered pair $(L_{\alpha,\beta}, \oplus_\beta)$ in a meaningful LFD system $(\mathcal{L}, \mathcal{O})$.

Proof Without loss of generality, we can assume that the ordered pair (L, \oplus) of initial code L is an abstract LFD-pair, and so satisfies the five conditions of Definition 9.25. By meaningfulness, we have:

$$L_{\alpha,\beta}(\lambda, v) = \alpha L\left(\frac{\lambda}{\alpha}, \frac{v}{\beta}\right) \quad \text{and} \quad v \oplus_\beta w = \beta\left(\frac{v}{\beta} \oplus \frac{w}{\beta}\right).$$

Conditions 1 to 4 readily follow. Condition 1 holds because, successively:

$$\begin{aligned} L_{\alpha,\beta}(\lambda, v) &\leq L_{\alpha,\beta}(\lambda', v') \\ &\iff \alpha L\left(\frac{\lambda}{\alpha}, \frac{v}{\beta}\right) \leq \alpha L\left(\frac{\lambda'}{\alpha}, \frac{v'}{\beta}\right) \quad (\text{by ST-meaningfulness}) \\ &\iff \alpha L\left(a \frac{\lambda}{\alpha}, \frac{v}{\beta}\right) \leq \alpha L\left(a \frac{\lambda'}{\alpha}, \frac{v'}{\beta}\right) \quad (\text{by Condition 1 for } (L, \oplus)) \\ &\iff L_{\alpha,\beta}(a\lambda, v) \leq L_{\alpha,\beta}(a\lambda', v') \quad (\text{by ST-meaningfulness}). \end{aligned}$$

For Condition 3, we have $\lim_{v \rightarrow c} L_{\alpha,\beta}(\lambda, v) = \alpha \lim_{\frac{v}{\beta} \rightarrow \frac{c}{\beta}} L\left(\frac{\lambda}{\alpha}, \frac{v}{\beta}\right) = 0$ (cf. Remark 9.43). We omit the proofs of Conditions 2 and 4 which are straightforward consequences of ST-meaningfulness.

We turn to Condition 5. Since Axioms [R] and [M] are equivalent by Lemma 9.26, it suffices to prove that the ordered pair $(L_{\alpha,\beta}, \oplus_\beta)$ satisfies Axiom [R]. By the ST-meaningfulness of \mathcal{L} ,

$$L_{\alpha,\beta}(L_{\alpha,\beta}(\lambda, v), w) = \alpha L\left(\frac{L_{\alpha,\beta}(\lambda, v)}{\alpha}, \frac{w}{\beta}\right) = \alpha L\left(\frac{\alpha L\left(\frac{\lambda}{\alpha}, \frac{v}{\beta}\right)}{\alpha}, \frac{w}{\beta}\right).$$

Cancelling the α 's in the fraction inside the parentheses in the r.h.s. gives

$$\begin{aligned}
 L_{\alpha,\beta}(L_{\alpha,\beta}(\lambda, v), w) &= \alpha L\left(L\left(\frac{\lambda}{\alpha}, \frac{v}{\beta}\right), \frac{w}{\beta}\right) \\
 &= \alpha L\left(\frac{\lambda}{\alpha}, \frac{v}{\beta} \oplus \frac{w}{\beta}\right) \quad \left(\begin{array}{l} \text{by Axiom [R]} \\ \text{applied to } L \end{array}\right) \\
 &= \alpha L\left(\frac{\lambda}{\alpha}, \frac{1}{\beta}(v \oplus_\beta w)\right) \quad (\text{meaningfulness of } \mathcal{O}) \\
 &= L_{\alpha,\beta}(\lambda, v \oplus_\beta w) \quad (\text{ST-meaningfulness of } \mathcal{L}).
 \end{aligned}$$

□

9.11 Meaningful Representations of Scientific Ratio-scaled Codes

This section and the next contain results that have been previewed earlier in Part II. Here in Section 9.11, which recalls results of Falmagne (2014, 2015) and Falmagne and Doble (2015), we present some consequences of pairing the meaningfulness condition with the abstract axioms of associativity, quasi-permutability, bisymmetry, translatability, and the LFD Axiom, the latter in the context of LFD systems. We include only a couple of exemplary proofs; for the remaining proofs, see chapter 7 in Falmagne and Doble (2015). This section otherwise follows closely Falmagne and Doble (2015).

9.11.1 Associativity and the Pythagorean Theorem

Theorem 9.45 Suppose that $\mathcal{F} = \{F_\alpha \mid \alpha \in \mathbb{R}_{++}\}$ is a $(\frac{1}{2}, \frac{1}{2})$ -meaningful ST-collection of codes, with $F_\alpha : \mathbb{R}_{++} \times \mathbb{R}_{++} \xrightarrow{\text{onto}} \mathbb{R}_{++}$ for all α in \mathbb{R}_{++} . If one of these codes is homogeneous and associative, then any code $F_\alpha \in \mathcal{F}$ must have, for some positive constant θ , the form

$$F_\alpha(x, y) = (x^\theta + y^\theta)^{\frac{1}{\theta}} = F(x, y). \quad (9.85)$$

This result is essentially the Pythagorean theorem, up to the exponent. We know that the hypotenuse of a right triangle is an associative function of the length of the two sides, as detailed in Figure 9.1. The fact that θ equals 2 can be derived from the area of the square postulate and a couple of other intuitively obvious postulates of geometry, as follows.

Remark 9.46 Consider the three following postulates of geometry.

1. **The area of the square postulate.** If the side of a square has length x , then the area is equal to x^2 .
2. **Area addition postulate.** The area of a region is the sum of the areas of its non-overlapping parts.
3. **Congruence postulate.** If two polygons are congruent, then they have the same area.

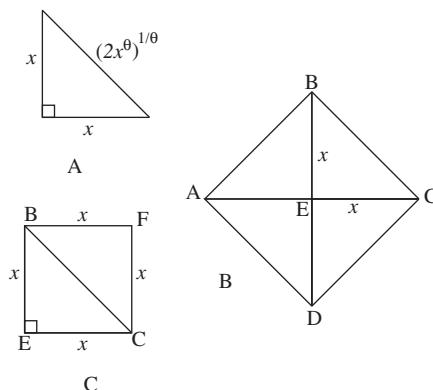


Figure 9.2 Construction of the four congruent triangles forming a square in the completion of the proof of the Pythagorean theorem.

If we add these three postulates²⁰ to the conditions of Theorem 9.45, we can derive the Pythagorean theorem, that is, we obtain $\theta = 2$ in Equation (9.85). Here is the argument, which is based on Figure 9.2.

Draw a right triangle with both leg lengths equal to x (see Figure 9.2A). Make three copies of the triangle, giving four congruent isosceles triangles. Each of the triangles has leg lengths of x , equal base angles, and by Theorem 9.45, a hypotenuse of length $(x^\theta + x^\theta)^{1/\theta} = (2x^\theta)^{1/\theta}$. Arrange the four triangles to form square ABCD, as in Figure 9.2B. By construction, this square has side lengths $(2x^\theta)^{1/\theta}$. The area of the square postulate implies that the area of ABCD is $((2x^\theta)^{1/\theta})^2$. The area of each of the four triangles is equal to $\frac{1}{2}x^2$. (For example, the area of the triangle BEC is equal to half of the area of the square BECF, which is equal to x^2 by the area of the square postulate; see Figure 9.2C.) So the area of the square ABCD must be equal to $4 \times \frac{1}{2}x^2 = 2x^2$. The two expressions for the area of the square must be equal, that is, we must have

$$2x^2 = ((2x^\theta)^{1/\theta})^2 = 2^{\frac{2}{\theta}}x^2, \quad \text{which yields } \theta = 2.$$

9.11.2 Meaningful Quasi-permutable Laws

The following theorem, which describes some results when quasi-permutability combines with meaningfulness, is in a similar spirit to that of Theorem 9.45 above. Interestingly, the following theorem provides yet another proof of the Pythagorean theorem, up to the exponent: we show in Subsection 9.11.3 that the family of codes measuring the hypotenuse of a right triangle with leg lengths x and y satisfies the type of quasi-permutability called for by the theorem, and so, applying Equation (9.93) below, we get that the initial code P is of the form $P(x, y) = (x^\theta + y^\theta)^{1/\theta}$.

20 Along with the fact that the sum of the angles of a triangle is equal to 180° .

Theorem 9.47 Let $\mathcal{F} = \{F_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ be an ST-meaningful collection of 2-codes, with $F_{\alpha,\beta} : \mathbb{R}_{++} \times \mathbb{R}_{++} \xrightarrow{\text{onto}} \mathbb{R}_{++}$ for all $\alpha, \beta \in \mathbb{R}_{++}$. Moreover, suppose that each code is permutable with respect to the initial code $F = F_{1,1}$ and that one of the codes is strictly increasing in both variables. Then there are constants $\phi, \gamma > 0$ such that Case A holds:

$$\text{Case A.} \quad F(y, r) = \phi y r^\gamma, \quad (9.86)$$

$$F_{\alpha,\beta}(y, r) = \phi y \left(\frac{r}{\beta}\right)^\gamma \quad \text{for all } F_{\alpha,\beta} \in \mathcal{F}. \quad (9.87)$$

If $F_{\alpha,\beta} = F_{\alpha,\alpha} = F_\alpha$ for each code in \mathcal{F} , so that $\mathcal{F} = \{F_\alpha \mid \alpha \in \mathbb{R}_{++}\}$, with all of the other hypotheses the same as above, then for some positive constants ϕ, γ, θ and λ and some constant η , either Case B or Case C holds:

$$\text{Case B.} \quad F(y, r) = \phi y r^\gamma, \quad (9.88)$$

$$F_\alpha(y, r) = \phi y \left(\frac{r}{\alpha}\right)^\gamma \quad \text{for all } F_\alpha \in \mathcal{F}. \quad (9.89)$$

If some code in \mathcal{F} is symmetric, then $\gamma = 1$, and if in addition some code is homogeneous, then

$$F_\alpha(y, r) = F(y, r) = \theta y r \quad \text{for all } F_\alpha \in \mathcal{F}. \quad (9.90)$$

$$\text{Case C.} \quad F(y, r) = (y^\theta + \lambda r^\theta + \eta)^{\frac{1}{\theta}}, \quad (9.91)$$

$$F_\alpha(y, r) = (y^\theta + \lambda r^\theta + \alpha^\theta \eta)^{\frac{1}{\theta}} \quad \text{for all } F_\alpha \in \mathcal{F}. \quad (9.92)$$

If some code in \mathcal{F} is symmetric, then $\lambda = 1$, and if in addition some code is homogeneous, then

$$F_\alpha(y, r) = F(y, r) = (y^\theta + r^\theta)^{\frac{1}{\theta}} \quad \text{for all } F_\alpha \in \mathcal{F}. \quad (9.93)$$

Proof **Case A.** Since each code in the family is permutable with respect to the initial code F , by Lemma 9.16, the code F is permutable. Using Lemma 9.17(ii) and the fact that \mathcal{F} is a ST-meaningful collection, we get for all $F_{\alpha,\beta}$ in \mathcal{F} :

$$F_{\alpha,\beta}(y, r) = \alpha F\left(\frac{y}{\alpha}, \frac{r}{\beta}\right) = \alpha f^{-1}\left(f\left(\frac{y}{\alpha}\right) + g\left(\frac{r}{\beta}\right)\right), \quad (9.94)$$

for some continuous, strictly increasing functions f and g , with in particular

$$F(y, r) = f^{-1}(f(y) + g(r)). \quad (9.95)$$

□

Using Lemma 9.17(ii) (twice) and meaningfulness, we get successively

$$\begin{aligned} F_{\alpha,\beta}(F(y, r), z) \\ = F_{\alpha,\beta}(f^{-1}(f(y) + g(r)), z) \end{aligned} \quad (9.96)$$

$$= \alpha F\left(\frac{1}{\alpha}f^{-1}(f(y) + g(r)), \frac{z}{\beta}\right) \quad (9.97)$$

$$= \alpha f^{-1}\left(f\left(\frac{1}{\alpha}f^{-1}(f(y) + g(r))\right) + g\left(\frac{z}{\beta}\right)\right) \quad (9.98)$$

$$= \alpha f^{-1}\left(f\left(\frac{1}{\alpha}f^{-1}(f(y) + g(z))\right) + g\left(\frac{r}{\beta}\right)\right) \quad (9.99)$$

by quasi-permutability.

Equating the last two right-hand sides, canceling the α 's, and applying the function f on both sides, we get

$$f\left(\frac{1}{\alpha}f^{-1}(f(y) + g(r))\right) + g\left(\frac{z}{\beta}\right) = f\left(\frac{1}{\alpha}f^{-1}(f(y) + g(z))\right) + g\left(\frac{r}{\beta}\right). \quad (9.100)$$

Setting $s = f(y)$, $t = g(r)$, fixing $z = 1$, and temporarily assuming that $\frac{1}{\alpha} = \frac{1}{\beta} = v$, Equation (9.100) becomes

$$f\left(vf^{-1}(s+t)\right) + g(v) = f\left(vf^{-1}(s+g(1))\right) + g\left(vg^{-1}(t)\right). \quad (9.101)$$

Defining the functions

$$h_v = f \circ vf^{-1}, \quad m_v : s \mapsto f\left(vf^{-1}(s+g(1))\right), \quad k_v : t \mapsto g\left(vg^{-1}(t)\right) - g(v),$$

Equation (9.101) becomes

$$h_v(s+t) = m_v(s) + k_v(t),$$

a Pexider equation. (See, e.g., section 3.1 of Aczél 1966.) In view of the conditions on the functions, the solution is

$$h_v(s) = p(v)s + q(v) + w(v) \quad (9.102)$$

$$m_v(s) = p(v)s + q(v)$$

$$k_v(t) = p(v)t + w(v), \quad (9.103)$$

for some constants $p(v)$, $q(v)$ and $w(v)$ possibly varying with v . Rewriting (9.102) and (9.103) in terms of the functions f and g , we obtain, with $\zeta(v) = q(v) + w(v)$,

$$\begin{aligned} h_v(s) &= (f \circ vf^{-1})(s) = p(v)s + \zeta(v), \\ k_v(t) &= g\left(vg^{-1}(t)\right) - g(v) = p(v)t + w(v) \end{aligned} \quad (9.104)$$

yielding, with $\gamma(v) = w(v) + g(v)$,

$$f(vy) = p(v)f(y) + \zeta(v), \quad (9.105)$$

$$g(vr) = p(v)g(r) + \gamma(v). \quad (9.106)$$

These are Vincze equations (cf. Vincze, 1962). In principle, for each of Equations (9.105) and (9.106), we have two solutions for the functions f and g depending on whether or not $p(v)$ is a constant function. But only the case below is consistent with the hypotheses.

Suppose that p is a constant function. We have then, for some positive constants b and d and some constants a and c ,

$$f(y) = b \ln y + a, \quad \text{and so} \quad f^{-1}(z) = e^{\frac{z-a}{b}} \quad (9.107)$$

$$g(r) = d \ln r + c. \quad (9.108)$$

Rewriting F in terms of the solutions (9.107) and (9.108) for the functions f and g yields

$$\begin{aligned} F(y, r) &= f^{-1}(f(y) + g(r)) = e^{\frac{f(y)+g(r)-a}{b}} \\ &= e^{\frac{b \ln y + a + d \ln r + c - a}{b}} = y e^{\frac{d \ln r + c}{b}} = e^{\frac{c}{b}} y r^{\frac{d}{b}}, \end{aligned}$$

and with $\phi = e^{\frac{c}{b}}$ and $\gamma = \frac{d}{b}$,

$$F(y, r) = \phi y r^\gamma. \quad (9.109)$$

With $F_{\alpha, \beta}(y, r) = \alpha F\left(\frac{y}{\alpha}, \frac{r}{\beta}\right)$, we get

$$F_{\alpha, \beta}(y, r) = \phi y \left(\frac{r}{\beta}\right)^\gamma. \quad (9.110)$$

It is easily verified that (9.109) and (9.110) imply permutability with respect to F :

$$\begin{aligned} F_{\alpha, \beta}(F(y, r), t) &= \phi F(y, r) \left(\frac{t}{\beta}\right)^\gamma = \phi^2 y r^\gamma \left(\frac{t}{\beta}\right)^\gamma = \phi^2 y \left(\frac{1}{\beta}\right)^\gamma r^\gamma t^\gamma \\ &= F_{\alpha, \beta}(F(y, t), r). \end{aligned}$$

Observation. If p takes at least two distinct values, then the form obtained for the functions $F_{\alpha, \beta}$ results in a collection \mathcal{F} that is not permutable with respect to F . The argument goes as follows. From (9.105) and (9.106) we get, for some $\theta > 0$:

$$f(y) = b y^\theta + a \quad \text{for some constants } b > 0 \text{ and } a, \quad (9.111)$$

$$g(r) = d r^\theta + c \quad \text{for some constants } d > 0 \text{ and } c. \quad (9.112)$$

From (9.111), we obtain

$$f^{-1}(t) = \left(\frac{t-a}{b}\right)^{\frac{1}{\theta}}. \quad (9.113)$$

Computing $F(y, r)$, we obtain from (9.111), (9.113) and (9.112), successively

$$\begin{aligned} F(y, r) &= f^{-1}(f(y) + g(r)) = \left(\frac{f(y) + g(r) - a}{b}\right)^{\frac{1}{\theta}} \\ &= \left(\frac{b y^\theta + a + d r^\theta + c - a}{b}\right)^{\frac{1}{\theta}} = \left(y^\theta + \frac{d}{b} r^\theta + \frac{c}{b}\right)^{\frac{1}{\theta}} \end{aligned}$$

and with $\lambda = \frac{d}{b}$ and $\eta = \frac{c}{b}$, finally

$$F(y, r) = (y^\theta + \lambda r^\theta + \eta)^{\frac{1}{\theta}}.$$

With meaningfulness, this gives for any $\alpha > 0$ and $\beta > 0$

$$F_{\alpha, \beta}(y, r) = \alpha \left(\left(\frac{y}{\alpha} \right)^\theta + \lambda \left(\frac{r}{\beta} \right)^\theta + \eta \right)^{\frac{1}{\theta}} = \left(y^\theta + \lambda \left(\frac{\alpha r}{\beta} \right)^\theta + \alpha^\theta \eta \right)^{\frac{1}{\theta}}.$$

Since we have:

$$F_{\alpha, \beta}(F(y, r), t) = \left(y^\theta + \lambda r^\theta + \eta + \lambda \left(\frac{\alpha t}{\beta} \right)^\theta + \alpha^\theta \eta \right)^{\frac{1}{\theta}},$$

assuming that $F_{\alpha, \beta}(F(y, r), t) = F_{\alpha, \beta}(F(y, t), r)$ leads to

$$\left(y^\theta + \lambda r^\theta + \eta + \lambda \left(\frac{\alpha t}{\beta} \right)^\theta + \alpha^\theta \eta \right)^{\frac{1}{\theta}} = \left(y^\theta + \lambda t^\theta + \eta + \lambda \left(\frac{\alpha r}{\beta} \right)^\theta + \alpha^\theta \eta \right)^{\frac{1}{\theta}},$$

and after simplification we get

$$r^\theta + \left(\frac{\alpha t}{\beta} \right)^\theta = t^\theta + \left(\frac{\alpha r}{\beta} \right)^\theta. \quad (9.114)$$

For (9.114) to hold for all $\alpha, \beta > 0$, it must be that $r^\theta = t^\theta$.

This completes our proof of Case A.

Notes on Cases B and C. We now assume that $F_{\alpha, \beta} = F_{\alpha, \alpha} = F_\alpha$ for each code in \mathcal{F} , so that $\mathcal{F} = \{F_\alpha \mid \alpha \in \mathbb{R}_{++}\}$. Applying the same derivation as in Case A, except this time assuming $\alpha = \beta$, we again obtain Equations (9.105) and (9.106):

$$f(vy) = p(v)f(y) + \zeta(v),$$

$$g(vr) = p(v)g(r) + \gamma(v).$$

(Note that there is no assumption that α and β vary independently in obtaining Equations (9.105) and (9.106). For example, in going from (9.100) to (9.101), it is assumed that $\frac{1}{\alpha} = \frac{1}{\beta} = v$.) Again we have two cases, determined by whether or not $p(v)$ is constant.

Suppose that p is a constant function. (This gives Case B of the theorem.) We again obtain Equation (9.109),

$$F(y, r) = \phi yr^\gamma,$$

and meaningfulness then gives, for all $\alpha > 0$,

$$F_\alpha(y, r) = \alpha F\left(\frac{y}{\alpha}, \frac{r}{\alpha}\right) = \phi y \left(\frac{r}{\alpha}\right)^\gamma. \quad (9.115)$$

If some code in \mathcal{F} is symmetric, then by Lemma 9.40(iv), all the codes in \mathcal{F} are symmetric. Then necessarily $\gamma = 1$, that is,

$$F(y, r) = \phi yr. \quad (9.116)$$

If in addition some code in \mathcal{F} is homogeneous, then by Lemma 9.41(iv), all the codes are homogeneous, and furthermore $F_\alpha(y, r) = F(y, r)$ for all $\alpha > 0$. This gives

$$F_\alpha(y, r) = \phi y r, \quad (9.117)$$

which completes the proof of Case B.

Now suppose that p takes at least two distinct values. (Case C.)

Proceeding as previously, we again get Equations (9.111)–(9.113), which result in

$$F(y, r) = (y^\theta + \lambda r^\theta + \eta)^{\frac{1}{\theta}} \quad (9.118)$$

for some positive constants θ and λ , and for some constant η .

From meaningfulness, this gives for all $\alpha > 0$,

$$F_\alpha(y, r) = \alpha \left(\left(\frac{y}{\alpha} \right)^\theta + \lambda \left(\frac{r}{\alpha} \right)^\theta + \eta \right)^{\frac{1}{\theta}} \quad (9.119)$$

$$= (y^\theta + \lambda r^\theta + \alpha^\theta \eta)^{\frac{1}{\theta}}. \quad (9.120)$$

The functions F and F_α in Equations (9.118) and (9.120) satisfy $F_\alpha(F(y, r), t) = F_\alpha(F(y, t), r)$ for all $\alpha, y, r, t > 0$. Now, if some code in \mathcal{F} is symmetric, then all the codes in \mathcal{F} are symmetric, and so $\lambda = 1$, that is,

$$F(y, r) = (y^\theta + r^\theta + \eta)^{\frac{1}{\theta}} \quad (9.121)$$

and

$$F_\alpha(y, r) = (y^\theta + r^\theta + \alpha^\theta \eta)^{\frac{1}{\theta}} \quad \text{for all } \alpha > 0. \quad (9.122)$$

If in addition some code in \mathcal{F} is homogeneous, then all the codes are. Moreover, Lemma 9.41 gives that $F_\alpha(y, r) = F(y, r)$ for all $\alpha > 0$. This means that $(y^\theta + r^\theta + \alpha^\theta \eta)^{\frac{1}{\theta}} = (y^\theta + r^\theta + \eta)^{\frac{1}{\theta}}$ for all $\alpha > 0$, which implies $\eta = 0$. We thus have $F_\alpha(y, r) = F(y, r) = (y^\theta + r^\theta)^{\frac{1}{\theta}}$. This completes the proof of Case C. \square

9.11.3 Quasi-permutability and the Pythagorean Theorem

As previewed at the beginning of Subsection 9.11.2, we get another proof of the Pythagorean theorem via Equation (9.93) of Theorem 9.47.²¹ What remains in order to conclude the form in Equation (9.93) is to show that the quasi-permutability hypothesis of the theorem is satisfied.

To this end, we suppose that the length $P(x, y)$ of the hypotenuse of a right triangle with leg lengths x and y is a symmetric, homogeneous code. We take the function P to be the initial code of a family of codes $\{P_\alpha\}$ and establish the permutability of P and the quasi-permutability of the code P_α (its permutability with respect to P), for any $\alpha > 0$, by an elementary geometric argument. This argument

²¹ The fact that $\theta = 2$ in Equation (9.93) follows from the argument in Remark 9.46.

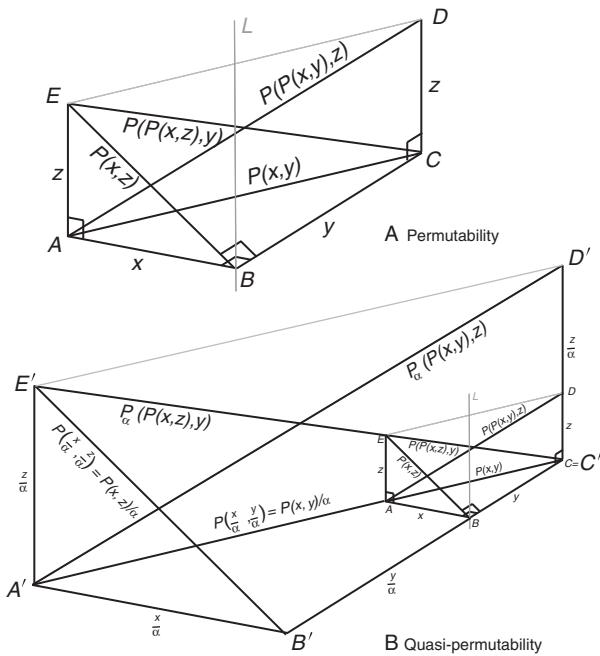


Figure 9.3 The upper graph, A, illustrates the permutability of P formalized by the equation $P(P(x,y),z) = P(P(x,z),y)$. The lower graph, B, shows that the quasi-permutability condition formalized by the equation $P_\alpha(P(x,y),z) = P_\alpha(P(x,z),y)$ only involves a rescaling of all the variables pictured in Figure 9.3A, resulting in a similar figure, with the rectangle $A'C'D'E'$ similar to the rectangle $ACDE$. The measures of the two diagonals of the rectangle $A'C'D'E'$ are $P_\alpha(P(x,y),z)$ and $P_\alpha(P(x,z),y)$.

is drawn in Figure 9.3. Note that the quasi-permutability condition is related to the similarity of the two figures $ABCDE$ and $A'B'C'D'E'$.

A right triangle $\triangle ABC$ with leg lengths x and y and hypotenuse of length $P(x,y)$ is pictured in the lower part of Figure 9.3A. Thus $AB = x$, $BC = y$ and $P(x,y) = AC$. Another right triangle $\triangle ACD$ is defined by the segment \overline{CD} of length z , which is perpendicular to the plane of $\triangle ABC$.

The length of the hypotenuse \overline{AD} of $\triangle ACD$ is thus $P(P(x,y),z) = AD$. Still another right triangle $\triangle EAB$ is defined by the perpendicular \overline{AE} to the plane of $\triangle ABC$. We choose E such that $AE = z = CD$; we have thus $EB = P(x,z)$. Since \overline{AE} is perpendicular to the plane of the right triangle $\triangle ABC$, \overline{EB} is perpendicular to \overline{BC} . The lines \overline{BC} and \overline{BE} are perpendicular. (Indeed, the perpendicular L at the point B to the plane of triangle $\triangle ABC$ is coplanar with \overline{AE} . So, as \overline{BC} is perpendicular to both \overline{AE} and L , it must be perpendicular to the plane of $\triangle EAB$, and so it must be perpendicular to \overline{EB} .) Accordingly, $EC = P(P(x,z),y)$ is the length of the hypotenuse of the right triangle $\triangle EBC$. It is clear that, by construction, the four points A , C , D and E are coplanar. They define a rectangle whose diagonals \overline{AD}

and \overline{EC} must be equal. So, we must have $P(P(x, y), z) = P(P(x, z), y)$, establishing the permutability of the code P .

The quasi-permutability of P_α . For any positive real number α , the triangle $\triangle A'B'C'$ pictured in Figure 9.3B, with $C' = C$, A collinear with $A'C'$, B collinear with $B'C'$, and $A'B' = \frac{x}{\alpha}$, $B'C' = \frac{y}{\alpha}$ and $A'C' = \frac{P(x, y)}{\alpha}$, is similar to the triangle $\triangle ABC$ also represented in Figure 9.3B. So, we have

$$P\left(\frac{x}{\alpha}, \frac{y}{\alpha}\right) = \frac{P(x, y)}{\alpha}. \quad (9.123)$$

The function P is the initial code of the meaningful family of codes $\{P_\alpha\}$. For the code P_α in that family, we get

$$\begin{aligned} P_\alpha(P(x, y), z) &= \alpha P\left(\frac{P(x, y)}{\alpha}, \frac{z}{\alpha}\right) && \text{(by meaningfulness)} \\ &= \alpha P\left(P\left(\frac{x}{\alpha}, \frac{y}{\alpha}\right), \frac{z}{\alpha}\right) && \text{(by Equation (9.123))} \\ &= \alpha P\left(P\left(\frac{x}{\alpha}, \frac{z}{\alpha}\right), \frac{y}{\alpha}\right) && \text{(by the permutability of } P\text{)} \\ &= \alpha P\left(\frac{P(x, z)}{\alpha}, \frac{y}{\alpha}\right) && \text{(by Equation (9.123))} \\ &= P_\alpha(P(x, z), y) && \text{(by meaningfulness).} \end{aligned}$$

We conclude that any code P_α in the family $\{P_\alpha\}$ is quasi-permutable, in particular, is permutable with respect to the initial code P .

9.11.4 Meaningful Bisymmetric Laws

Theorem 9.48 *Let $\mathcal{F} = \{F_\alpha \mid \alpha \in \mathbb{R}_{++}\}$ be an ST-meaningful collection of codes, with $F_\alpha : J \times J \xrightarrow{\text{onto}} J$ for all α in \mathbb{R}_{++} , with J a closed interval.²² Moreover, suppose that one of these codes, say the code F_β , is strictly increasing in both variables, homogeneous, bisymmetric, and moreover satisfies $F_\beta(x, x) = x$ for all $x \in \mathbb{R}_+$. Then, we have two possible forms for the codes in the collection \mathcal{F} :*

1. $F_\alpha(x, y) = x^{1-q}y^q$ *for some $q \neq 0, 1$, and all codes $F_\alpha \in \mathcal{F}$;*
2. $F_\alpha(x, y) = ((1 - q)x^\theta + qy^\theta)^{\frac{1}{\theta}}$ *for some $\theta > 0$, some $q \neq 0, 1$, and all codes $F_\alpha \in \mathcal{F}$.*

So, if the function F_β is a symmetric function, we have either:

3. $F_\alpha(x, y) = (xy)^{\frac{1}{2}}$ *for all codes $F_\alpha \in \mathcal{F}$;*

or

4. $F_\alpha(x, y) = \left(\frac{x^\theta + y^\theta}{2}\right)^{\frac{1}{\theta}}$ *for some $\theta > 0$ and all codes $F_\alpha \in \mathcal{F}$.*

²² Note that, in Cases 1, 2, 3 and 4, the formula in the r.h.s. does not vary with the unit. So, the assumption $F_\alpha : J \times J \xrightarrow{\text{onto}} J$ for all α in \mathbb{R}_{++} makes sense.

9.11.5 Meaningful Translatable Laws

The following theorem, previewed earlier, states that the exponential decay model given by Beer's Law is the only functional form that can arise for meaningful families of codes that satisfy translatability and left homogeneity.

Theorem 9.49 *Let $\mathcal{F} = \{F_{\mu,v} \mid \mu, v \in \mathbb{R}_{++}\}$ be an ST-meaningful collection of codes, with $F_{\mu,v} : \mathbb{R}_{++} \times \mathbb{R}_{++} \xrightarrow{\text{onto}} \mathbb{R}_{++}$. Suppose that one of these codes, say the code $F_{\alpha,\beta}$, is strictly decreasing in the second variable, translatable, and left homogeneous of degree one, that is: for any a in \mathbb{R}_{++} , we have $F_{\alpha,\beta}(ax, y) = aF_{\alpha,\beta}(x, y)$. Then there must be a positive constant c such that the initial code F has the form*

$$F(x, y) = xe^{-\frac{y}{c}} \quad (9.124)$$

and for any code $F_{\mu,v} \in \mathcal{F}$

$$F_{\mu,v}(x, y) = xe^{-\frac{y}{vc}}. \quad (9.125)$$

9.12 Order-invariance under Transformations

Sections 9.9 through 9.11 examined the effect of the invariance conditions given by Definition 9.27 (ratio scale meaningfulness) and Definition 9.32 (1-1-meaningfulness and order-meaningfulness) on the mathematical form of (families of) numerical codes giving scientific or mathematical laws. This examination included the consequences of some abstract axioms such as associativity, quasi-permutability, and translatability. The current section is in a similar spirit, with the abstract axioms replaced by “order-invariance” axioms. An example of the latter was given in Subsection 9.7.4 (see Definition 9.63 for the exact statement).

In the next subsection, we introduce three types of families of transformations. The first type is headed by the similarity transformations $x \mapsto \lambda x$. The second type is marked by convex transformations and involves functions with a common bounded domain, such as the closed interval $[0, c]$ of the LFD systems. The third type has families that are all related to the translation transformations, either of the form $x \mapsto x + \lambda$ or of the form $x \mapsto x + \lambda c$. (The difference is significant because c varies with the measurement unit.)

We describe in Subsection 9.12.2 the concept of a “transformation class”, which gathers families of transformations. Then, in Subsection 9.12.3, we give a number of preparatory results describing the effects of the transformation families and classes acting on a numerical code.

The remaining subsections contain the applications at the focus of our study. In Subsection 9.12.4, we apply the concept of meaningfulness to transformation families and classes. Subsection 9.12.5 describes the derivation of a generalized form of the LFD systems from four axioms involving meaningfulness and “order-invariance” concepts. In Subsections 9.12.6 and 9.12.7, we give axiomatizations of Beer's law and the monomial laws along the same lines. Not all scientific laws can

be analyzed by the methods of this section, as we demonstrate with van der Waals' equation.

This section follows closely Falmagne (2004) and chapter 8 of Falmagne and Doble (2015).

9.12.1 Transformation Families

Definition 9.50 Let J be a real interval, and suppose that \mathcal{F} is a family of continuous, strictly increasing functions $f : J \rightarrow J$, closed under function composition; then \mathcal{F} is a *transformation family* or, more briefly, a *T-family*, on J . The set J is the *domain* of the T-family \mathcal{F} . A T-family \mathcal{F} is *commutative* if $f \circ g = g \circ f$ for all $f, g \in \mathcal{F}$. It is *transitive* if, for every $x, y \in J$, there exists some $f \in \mathcal{F}$ such that $f(x) = y$.

Examples 9.51 (a) The T-family \mathcal{S} on \mathbb{R}_{++} containing, for all positive real numbers $\lambda \in \mathbb{R}_{++}$, the function $\sigma_\lambda : x \mapsto \lambda x$, is both commutative and transitive.

(b) With

$$p_\lambda :]0, 1[\rightarrow]0, 1[: x \mapsto \frac{\lambda x}{\lambda x + 1 - x}, \quad (\lambda \in \mathbb{R}_{++}), \quad (9.126)$$

the T-family $\mathcal{P} = \{p_\lambda \mid \lambda \in \mathbb{R}_{++}\}$ on $]0, 1[$ is also commutative and transitive.

- (c) For some $b, c \in [0, \infty[$ with $b < c$, let $\mathcal{C}_{[b,c]}$ be the T-family on $[b, c]$ containing, for any non-negative real numbers μ, λ satisfying $\mu + \lambda \leq 1$, the function $x \mapsto \mu b + \lambda x + (1 - \mu - \lambda)c$. This T-family is transitive but not commutative.
- (d) Two subfamilies of $\mathcal{C}_{[b,c]}$ are commutative but not transitive, namely: the T-family $\mathcal{C}_{[b,c]}^+$ containing all the functions $x \mapsto \lambda x + (1 - \lambda)c$, and the T-family $\mathcal{C}_{[b,c]}^-$ containing all the functions $x \mapsto \lambda b + (1 - \lambda)x$; we suppose that $0 < \lambda < 1$ in both of these cases.
- (e) For some fixed $c \in \mathbb{R}_{++}$, the T-family $\mathcal{T}_c = \{\tau_\lambda \mid \lambda \in \mathbb{R}_{++}\}$, with

$$\tau_\lambda : [0, \infty[\rightarrow [0, \infty[: x \mapsto x + \lambda c, \quad (9.127)$$

is also commutative and not transitive: for $(x, y) \in [0, \infty[^2$ with $y < x$, there is no $\lambda \in \mathbb{R}_{++}$ such that $\tau_\lambda(x) = y$.

Remark 9.52 Even though the T-family \mathcal{T}_c of Example 9.51(e) coincides with the T-family containing all the translations $x \mapsto x + \lambda$, $\lambda \in \mathbb{R}_{++}$, we distinguish between them because the constant c in the function $x \mapsto x + \lambda c$ is modified by any change of the measurement unit. This constant plays a critical role in two representation theorems (see Theorems 9.72 and 9.84).

Definition 9.53 The T-family \mathcal{S} on \mathbb{R}_{++} of Example 9.51(a) is called the *similarity* family. The T-families $\mathcal{C}_{[b,c]}$, $\mathcal{C}_{[b,c]}^+$ and $\mathcal{C}_{[b,c]}^-$ of Examples 9.51(c) and (d) are the *convex*, *upper convex*, and *lower convex* T-family on $[b, c]$, respectively. The T-family \mathcal{T}_c of Example 9.51(e) is the *(positive) relative translation* family on $[0, \infty[$ or the *c-relative translation* family on $[0, \infty[$.

Remark 9.54 The T-family \mathcal{P} of Example 9.51(b) has a domain bounded on both sides, but the relationship between it and the families of Examples 9.51(c) and (d) is only superficial. The critical distinction to draw is between (a)–(b) on the one hand, and (c)–(d)–(e) on the other hand. In all the cases in Example 9.51, the domains (and co-domains) of the transformations contain values of a scientific variable expressed in a ratio scale unit. A change of unit has no effect in Example 9.51(a) because the domain is \mathbb{R}_{++} . We shall see that the *pertinent* domain is also \mathbb{R}_{++} in Example 9.51(b) because we can express any transformation $p_\lambda \in \mathcal{P}$ as the composition

$$p_\lambda = u^{-1} \circ \sigma_\lambda \circ u :]0, 1[\xrightarrow{u} \mathbb{R}_{++} \xrightarrow{\sigma_\lambda} \mathbb{R}_{++} \xrightarrow{u^{-1}}]0, 1[\quad (9.128)$$

for some strictly increasing function u , with σ_λ in the similarity family \mathcal{S} of Example 9.51(a). See Definition 9.55 and Example 9.57(c) below.

In the cases of Examples 9.51(c)–(d), however, a change of unit means multiplying all quantities by some constant β , which shifts the domain of the family from $[b, c]$ to $[\beta b, \beta c]$, creating a new T-family of transformations. In the case of Example 9.51(e), the domain of the T-family does not change, but for any given $\lambda \in \mathbb{R}_{++}$, the transformation is modified by a change of unit. The concept of a family of transformations as specified by Definition 9.50 is not adequate to deal with these more complicated situations. We postpone the discussion of them for the moment (see Definition 9.60).

Definition 9.55 A T-family \mathcal{F} on J is *conjugate* to a T-family \mathcal{G} on J' if there exists some continuous, strictly increasing function u mapping J onto J' such that the function $g \mapsto u^{-1} \circ g \circ u$ is a 1-1 mapping of \mathcal{G} onto \mathcal{F} . In such a case, we say that \mathcal{F} is the *u-conjugate* of \mathcal{G} and we write $\mathcal{F} = u^{-1}\mathcal{G}u$. The function u will be called the *conjugation*.

Remarks 9.56 (a) Note that conjugation is an equivalence relation (symmetric, reflexive, and transitive). Families that are equivalent to the similarity family \mathcal{S} of Example 9.51(a) are given in Examples 9.57(a), (b), (c), and (d).
 (b) It is easily shown that a T-family which is conjugate to a commutative (resp. transitive) T-family is also commutative (resp., transitive).

Examples 9.57 (a) Let \mathcal{S} be the similarity family of Example 9.51(a), and suppose that u is the logarithmic function $u(x) = \ln x$. Then the T-family $u^{-1}\mathcal{S}u = \exp \mathcal{S} \ln$ is the ln-conjugate of \mathcal{S} and contains all the functions $x \mapsto \exp(\lambda \ln x) = x^\lambda$, with $\lambda \in \mathbb{R}_{++}$.
 (b) Again with the similarity family \mathcal{S} , suppose that $u(x) = e^x$. Then the family $u^{-1}\mathcal{S}u = \ln \mathcal{S} \exp$ is the exp-conjugate of \mathcal{S} and contains all the functions $x \mapsto x + \ln \lambda$, with $\lambda \in \mathbb{R}_{++}$.
 (c) The T-family \mathcal{S} is conjugate to the T-family $\mathcal{P} = \{p_\lambda \mid \lambda \in \mathbb{R}_{++}\}$ of Example 9.51(b). Specifically, take $u :]0, 1[\rightarrow \mathbb{R}_{++} : x \mapsto \frac{x}{1-x}$; then

$$p_\lambda(x) = (u^{-1} \circ (\lambda u))(x) = \frac{\lambda x}{\lambda x + 1 - x}. \quad (9.129)$$

- (d) Still another example of a T-family conjugate to \mathcal{S} is obtained by taking $u :]0, 1[\rightarrow \mathbb{R}_{++} : x \mapsto \ln \frac{x}{1-x}$, the logit function. The T-family $u^{-1}\mathcal{S}u$ contains all the transformations

$$(u^{-1} \circ (\lambda u))(x) = \frac{x^\lambda}{x^\lambda + (1-x)^\lambda}, \quad (0 < \lambda < 1). \quad (9.130)$$

As these examples show, there is a wide variety of families having essentially the same properties as those of the similarity family, namely, commutativity and transitivity. In fact, these two properties define the equivalence class of all the families conjugate to \mathcal{S} ; see Theorem 9.59 below. This theorem, along with results such as Theorem 9.70 and Lemma 9.79, highlights the importance of the similarity family in the foundation of scientific laws.

Definition 9.58 When a T-family \mathcal{F} is conjugate to the similarity family \mathcal{S} of Example 9.57(a), we say that \mathcal{F} is a *quasi-similarity* family. A T-family that is conjugate to the convex family $\mathcal{C}_{[b,c]}$ of Example 9.57(c) is called *quasi-convex*. The “quasi” prefix is also used for T-families that are conjugate to the upper convex, the lower convex, and the relative translation families. (Thus, we refer to such families as *quasi-upper convex*, etc.)

For instance, the families of Example 9.57(a), (b), (c) and (d) are all quasi-similarity families. Examples of quasi-convex (upper convex, lower convex) and quasi-relative translation families will be found later in this section.

The following is well known; see, e.g., Narens (2002a). (As will be seen, assuming meaningfulness allows for much weaker hypotheses for the family \mathcal{F} . Compare Lemma 9.79.)

Theorem 9.59 Any T-family \mathcal{F} on some interval J which is both transitive and commutative is a quasi-similarity family. Thus, there exists some continuous, strictly increasing function $u : J \rightarrow \mathbb{R}_{++}$ such that, for any $f \in \mathcal{F}$, there exists $\lambda \in \mathbb{R}_{++}$, with

$$f(x) = (u^{-1} \circ \lambda u)(x), \quad (x \in J). \quad (9.131)$$

The mapping $\lambda \mapsto u^{-1} \circ \lambda u = f$ is 1-1 from \mathbb{R}_{++} onto \mathcal{F} (cf. Definition 9.55).

As pointed out in Remark 9.54, the concept of a transformation family is wanting for certain cases of bounded domains because of the interplay between such domains and the measurement unit. The appropriate definitions are given in the next subsection. For simplicity, we restrict consideration to the convex, upper convex, lower convex, and relative translation families introduced in Examples 9.51(c)–(e) and Definition 9.53, and their conjugated transformation families.

9.12.2 Transformation Classes

Definition 9.60 For any $b, c \in [0, \infty[$ with $b < c$ and $\beta \in \mathbb{R}_{++}$, let $\mathcal{C}_{[\beta b, \beta c]}$ be the convex family on the interval $[\beta b, \beta c]$; thus, $\mathcal{C}_{[\beta b, \beta c]}$ contains all the functions

$$\xi_{\mu, \lambda; \beta} : [\beta b, \beta c] \rightarrow [\beta b, \beta c] : x \mapsto \mu \beta b + \lambda x + (1 - \mu - \lambda) \beta c, \quad (9.132)$$

$$(0 < \mu, 0 < \lambda, \mu + \lambda < 1).$$

The union $\Gamma_{[b, c]} = \cup_{\beta \in \mathbb{R}_{++}} \mathcal{C}_{[\beta b, \beta c]}$ is called the *convex transformation class* or, more briefly, the *convex T-class, of type [b, c]*. Similar definitions apply to the cases of the *upper convex* and *lower convex families*, yielding the *upper convex T-class* $\Gamma_{[b, c]}^+$ and the *lower convex T-class* $\Gamma_{[b, c]}^-$, both *of type [b, c]*, defined, respectively, by

$$\Gamma_{[b, c]}^+ = \cup_{\beta \in \mathbb{R}_{++}} \mathcal{C}_{[\beta b, \beta c]}^+, \quad \text{and} \quad \Gamma_{[b, c]}^- = \cup_{\beta \in \mathbb{R}_{++}} \mathcal{C}_{[\beta b, \beta c]}^-, \quad (9.133)$$

with (9.132) being replaced, respectively, by

$$\xi_{\lambda; \beta} : [\beta b, \beta c] \rightarrow [\beta b, \beta c] : x \mapsto \lambda x + (1 - \lambda) \beta c, \quad (0 < \lambda < 1) \quad (9.134)$$

and by

$$\xi_{\mu; \beta} : [\beta b, \beta c] \rightarrow [\beta b, \beta c] : x \mapsto \mu \beta b + (1 - \mu) x, \quad (0 < \mu < 1). \quad (9.135)$$

With $c \in \mathbb{R}_{++}$, the *c-relative translation T-class* is defined by the union $T_c = \cup_{\beta \in \mathbb{R}_{++}} T_{\beta c}$ which contains all the transformations

$$\tau_{\lambda; \beta} : [0, \infty[\rightarrow [0, \infty[: x \mapsto x + \lambda \beta c, \quad (\beta, \lambda \in \mathbb{R}_{++}). \quad (9.136)$$

The concept of conjugation previously defined for transformation families can be extended to T-classes.

Definition 9.61 Let $u : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ be any strictly increasing and continuous function. Then, for all $b, c \in [0, \infty[$ with $b < c$, and $\beta \in \mathbb{R}_{++}$, the restriction $u|_{[\beta b, \beta c]}$ of the function u to the closed interval $[\beta b, \beta c]$ maps this interval onto the closed interval $[u(\beta b), u(\beta c)]$. By abuse of notation, we write $u|_{[\beta b, \beta c]} = u$ for simplicity in what follows. (The domain of the restriction will always be clear from the context.) In such a situation, the family $u^{-1} \mathcal{C}_{[\beta b, \beta c]} u$ is *u-conjugate* to the convex family $\mathcal{C}_{[\beta b, \beta c]}$, and so is quasi-convex in the sense of Definition 9.58. The union

$$\Gamma_{[b, c]; u} = \cup_{\beta \in \mathbb{R}_{++}} u^{-1} \mathcal{C}_{[\beta b, \beta c]} u \quad (9.137)$$

is said to be *u-conjugate* to the convex T-class $\Gamma_{[b, c]}$ introduced in Definition 9.60. Extending our previous terminology and notation, we refer to $\Gamma_{[b, c]; u}$ as a *quasi-convex T-class of type [b, c]* and we write $\Gamma_{[b, c]; u} = u^{-1} \Gamma_{[b, c]} u$. Note that $\Gamma_{[b, c]; u}$ contains all the transformations

$$x \mapsto u^{-1} (\mu u(\beta b) + \lambda u(x) + (1 - \mu - \lambda) u(\beta c)) \quad (9.138)$$

$$(x \in [\beta b, \beta c], \beta \in \mathbb{R}_{++}, 0 < \mu, 0 < \lambda, \mu + \lambda < 1).$$

Similar conventions and notation apply in an obvious manner to the upper and lower convex cases, yielding *quasi-upper convex* and *quasi-lower convex T-classes* $\Gamma_{[b,c];u}^+ = u^{-1}\Gamma_{[b,c]}^+u$ and $\Gamma_{[b,c];u}^- = u^{-1}\Gamma_{[b,c]}^-u$ of type $[b, c]$, with (9.138) replaced, respectively, by

$$x \mapsto u^{-1}(\lambda u(x) + (1 - \lambda)u(\beta c)),$$

and

$$\begin{aligned} x &\mapsto u^{-1}(\mu u(\beta b) + (1 - \mu)u(x)) \\ (x &\in [\beta b, \beta c], \beta \in \mathbb{R}_{++}, 0 < \lambda < 1, 0 < \mu < 1). \end{aligned}$$

Such conventions and notation also apply to the relative translation case, giving the *quasi-c-relative translation T-class*

$$\mathbf{T}_{c;u} = u^{-1}\mathbf{T}_cu = \cup_{\beta \in \mathbb{R}_{++}}u^{-1}\mathcal{T}_{\beta c}u$$

which contains all the transformations

$$u^{-1} \circ \tau_{\lambda;\beta} \circ u : [0, \infty[\rightarrow [0, \infty[: x \mapsto u^{-1}(u(x) + \lambda u(\beta c)), \quad (\beta, \lambda \in \mathbb{R}_{++}).$$

Examples 9.62 (a) Suppose that u in (9.138) is the logarithmic function. Then $\Gamma_{[b,c];\ln} = \exp \Gamma_{[b,c]} \ln$ is the quasi-convex T-class containing all the functions

$$x \mapsto \exp(\mu \ln(\beta b) + \lambda \ln x + (1 - \mu - \lambda) \ln(\beta c)) = \left(\frac{b}{c}\right)^\mu x^\lambda (\beta c)^{1-\lambda}.$$

(b) The quasi-upper convex T-class $u\Gamma_{[b,c]}^+u^{-1}$ with $u(x) = ax^\theta + d$ and $a\theta > 0$ contains all the transformations

$$x \mapsto (\lambda x^\theta + (1 - \lambda)(\beta c)^\theta)^{1/\theta}, \quad (\beta \in \mathbb{R}_{++}, \theta \in \mathbb{R} \setminus \{0\}, 0 < \lambda < 1).$$

It will play an important role in our axiomatization of the LF-contraction.

9.12.3 Transformations Acting on Codes

Definition 9.63 A code $M : J \times J' \rightarrow \mathbb{R}_{++}$ is said to be *left order-invariant* with respect to a T-family \mathcal{F} on J , or simply *left order-invariant*, if for all $f \in \mathcal{F}$,

$$M(x, y) \leq M(z, w) \iff M(f(x), y) \leq M(f(z), w) \quad (x, z \in J, y, w \in J').$$

Right order-invariance is similarly defined for the second variable of the code M . We say that M is (*quasi-*) *similarity invariant* when M is either left or right order-invariant with respect to some (*quasi-*) similarity family, and similar language applies in the (*quasi-*) convex, upper convex, and lower convex cases, and in the (*quasi-*) relative translation case.

Lemma 9.64 A code M is left order-invariant with respect to some family \mathcal{F} if and only if there exists, for each $f \in \mathcal{F}$, some strictly increasing function $\phi_f : M(J \times J') \rightarrow M(J \times J')$ such that for all $x \in J$ and $y \in J'$, we have

$$(\phi_f \circ M)(x, y) = M(f(x), y). \tag{9.139}$$

The following condition is important in ensuring that a code has a “multiplicative representation” (as in Lemma 9.67 and Definition 9.68).

Definition 9.65 A code M satisfies *double cancellation* if, for all $x, z, t \in J$ and $y, w, s \in J'$, we have

$$M(x, y) \leq M(z, w) \& M(z, s) \leq M(t, y) \implies M(x, s) \leq M(t, w).$$

Lemma 9.66 A code M which is left order-invariant with respect to a commutative and transitive family necessarily satisfies double cancellation.

We omit the proof (see Falmagne & Doble, 2015, p. 112).

The following result is a reformulation, in the context of real variables, of a standard result in measurement theory (see, e.g., Krantz *et al.*, 1971). The proof is omitted.

Lemma 9.67 The two following conditions are equivalent for a numerical code M on $J \times J'$:

- (i) The code M satisfies double cancellation.
- (ii) There are two continuous functions $g : J \rightarrow \mathbb{R}_{++}$ and $h : J' \rightarrow \mathbb{R}_{++}$, respectively, comonotonic with the first and the second variable of M , and a continuous, strictly increasing function G defined for all values $g(x)h(y) \in \mathbb{R}_{++}$ with $(x, y) \in J \times J'$, such that $M(x, y) = G(g(x)h(y))$.

Definition 9.68 A code satisfying Condition (ii) in Lemma 9.67 is said to have a *multiplicative representation* (g, h, G) , with the functions g , h and G satisfying all the properties listed in Condition (ii). A code having a multiplicative representation is *multiplicative*.

The results in the remainder of this subsection – Lemma 9.69 and Theorems 9.70, 9.71 and 9.72 – pave the way for the representation theorems dealing with the LF-contraction (Theorem 9.82) and Beer's law (Theorem 9.84). For the proofs of the results in the rest of this subsection, the reader is referred to Falmagne and Doble (2015, pp. 113–119).

Lemma 9.69 Suppose that a multiplicative code M on $J \times J'$ is left order-invariant with respect to some family \mathcal{F} . Then, there is a multiplicative representation (g, h, G) such that for any $f \in \mathcal{F}$, the function ϕ_f in Equation (9.139) takes the form

$$\phi_f(s) = (G \circ g \circ f \circ g^{-1} \circ G^{-1})(s), \quad (9.140)$$

which defines ϕ_f in terms of the functions g , G and f . Accordingly, we obtain

$$(g \circ f \circ g^{-1})(g(x)h(y)) = g(f(x))h(y), \quad (f \in \mathcal{F}, x \in J, y \in J'). \quad (9.141)$$

The reader may be interested in comparing the following result to Theorems 9.38 and 9.39 in Subsection 9.9.4.

Theorem 9.70 Suppose that a code M on $J \times J'$ is left order-invariant with respect to some commutative and transitive family \mathcal{F} . Then M has a multiplicative representation. Specifically, there are two continuous functions $k : J \rightarrow \mathbb{R}_{++}$ and

$h : J' \rightarrow \mathbb{R}_{++}$, with k strictly increasing and h comonotonic with the second variable of M , a real number $\theta \neq 0$ such that k^θ is comonotonic with the first variable of M , and a continuous, strictly increasing function G defined for all values $k(x)^\theta h(y) \in \mathbb{R}_{++}$ with $(x, y) \in J \times J'$, such that

$$M(x, y) = G(k(x)^\theta h(y)). \quad (9.142)$$

Moreover, we have $k\mathcal{F}k^{-1} = \mathcal{S}$, where \mathcal{S} is the similarity family of Example 9.51(a) and Definition 9.53. In particular, if $J = \mathbb{R}_{++}$ and M is left order-invariant with respect to \mathcal{S} , then

$$M(x, y) = G(x^\theta h(y)). \quad (9.143)$$

Theorem 9.71(ii) below is the basic tool used in the proof of the representation Theorem 9.82, which deals with the LF-contraction. As shown in Theorem 9.82, requiring meaningfulness will result in specifying the function u in (9.144) as a power function.

Theorem 9.71 Suppose that M is a code on $J \times [0, c]$ having a multiplicative representation. Then, the following hold.

- (i) M is not right quasi-convex invariant.
- (ii) If M is right quasi-upper convex-invariant with $u : [0, c] \rightarrow [0, c]$ as the conjugation, then there is a continuous function $g : J \rightarrow \mathbb{R}_{++}$ comonotonic with the first variable of M , and a continuous, strictly increasing function G such that

$$M(x, y) = G\left(g(x)\left(1 - \frac{u(y) - u(0)}{u(c) - u(0)}\right)^\delta\right), \quad (x \in J, y \in [0, c]), \quad (9.144)$$

with $\delta < 0$ or $\delta > 0$ depending on whether M is increasing or decreasing in its second variable.

- (iii) If M is right quasi-lower convex-invariant with $u : [0, c] \rightarrow [0, c]$ as the conjugation satisfying $u(0) \neq 0$, then the functions g and G exist as above, but (9.144) is replaced by

$$M(x, y) = G\left(g(x)\left(1 - \frac{u(y)}{u(0)}\right)^\delta\right), \quad (x \in J, y \in [0, c]), \quad (9.145)$$

with the same constraints as in (ii) on the parameter δ .

What follows is a similar result leading to the representation theorem for Beer's law (Theorem 9.84).

Theorem 9.72 Let M be a code on $J \times [0, \infty[$, with M strictly decreasing in its second variable. Suppose also that M has a multiplicative representation and is right order-invariant with respect to a quasi- c -relative translation family, with $c \in \mathbb{R}_{++}$ (cf. Definitions 9.53 and 9.63); then there is a continuous function

$g : J \rightarrow [0, \infty[$ comonotonic with the first variable of M , a constant $\theta > 0$, and a continuous, strictly increasing function G such that

$$M(x, y) = G \left(g(x) \frac{\theta^{-\frac{u(y)}{u(c)} + 1}}{e} \right), \quad (x \in J, y \in [0, \infty[), \quad (9.146)$$

where $u : [0, \infty[\rightarrow [0, \infty[$ is the conjugation.

9.12.4 Meaningful Transformations

The meaningfulness condition given by Definition 9.27 applies to families of codes. The concept extends naturally to families of transformations, in the sense that the effects of the transformations must be consistent across changes of unit. We first consider the case of the T-families, in which the transformations are not indexed by the unit.

Definition 9.73 Let $c \in \mathbb{R}_{++}$ be a constant, and for each $\alpha, \beta \in \mathbb{R}_{++}$, let $L_{\alpha, \beta} : [0, \infty[\times [0, \beta c] \rightarrow [0, \infty[$ be a numerical code, which is strictly increasing in its first variable and strictly decreasing in its second variable. We write $L = L_{1,1}$ for the initial code and we suppose that the scale and the unit of $L_{\alpha, \beta}(\ell, v)$ and of ℓ are identical. The set $\mathcal{L}_c = \{L_{\alpha, \beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ of all such codes is referred to as a $\langle LF \rangle$ -collection.

Definition 9.74 A T-family \mathcal{F} is *left meaningful* with respect to a meaningful $\langle LF \rangle$ -collection $\mathcal{L}_c = \{L_{\alpha, \beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ if

$$\begin{aligned} \gamma L_{\alpha, \beta}(f(\alpha \ell), \beta v) &= \alpha L_{\gamma, v}(f(\gamma \ell), vv) \\ (\alpha, \beta, \gamma, v \in \mathbb{R}_{++}, 0 \leq \ell < \infty, 0 \leq v \leq c, f \in \mathcal{F}). \end{aligned}$$

Similar definitions of left meaningfulness for a T-family also apply to the collections relevant to Beer's law²³ and to those relevant to the monomial laws in two variables²⁴ (and in the latter case, also of right meaningfulness).

We now turn to T-classes. Here, a change of the unit of the second variable of the code, say from β to v , must be accompanied by the same change of unit for the transformations in the relevant T-class.

Definition 9.75 Let $\Gamma_{[0, c]; u}^+$ be the quasi-upper convex T-class of type $[0, c]$ with conjugation u ; thus $\Gamma_{[0, c]; u}^+$ contains all the transformations

$$\xi_{\lambda; \beta}(v) = u^{-1} (\lambda u(v) + (1 - \lambda)u(\beta c)) \quad (\beta \in \mathbb{R}_{++}, 0 < \lambda < 1). \quad (9.147)$$

The T-class $\Gamma_{[0, c]; u}^+$ is *right meaningful* with respect to the meaningful $\langle LF \rangle$ -collection of codes $\mathcal{L}_c = \{L_{\alpha, \beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ if

²³ See Subsection 9.12.6 and Theorem 9.84.

²⁴ See Subsection 9.12.7 and Theorem 9.86.

$$\begin{aligned} \gamma L_{\alpha,\beta}(\alpha\ell, \xi_{\lambda;\beta}(\beta v)) &= \alpha L_{\gamma,v}(\gamma\ell, \xi_{\lambda;v}(vv)) \\ (\alpha, \beta, \gamma, v \in \mathbb{R}_{++}, 0 \leq \ell < \infty, 0 \leq v \leq c, 0 < \lambda < 1). \end{aligned}$$

Similar definitions of right meaningfulness apply to the quasi-lower convex T-class $\Gamma_{[0,c];u}^-$, and also (with respect to the collection of codes \mathcal{I} relevant to Beer's law) to the quasi- c -relative translation T-class $\mathbf{T}_{c;u}$, which contains all the transformations

$$\eta_{\lambda;\beta} : [0, \infty[\rightarrow [0, \infty[: y \mapsto u^{-1}(u(y) + \lambda u(\beta c)), \quad (\beta, \lambda \in \mathbb{R}_{++}).$$

To distinguish the meaningfulness of a collection of codes from that of a T-family or a T-class with respect to a collection, we refer to the former as *C-meaningfulness* and to the latter as *T-meaningfulness*. The two concepts of meaningfulness are related: for instance, if \mathcal{L}_c is meaningful, then the similarity family \mathcal{S} (Definition 9.53) is left meaningful with respect to \mathcal{L}_c .

As shown in the next three lemmas, combining C-meaningfulness and T-meaningfulness has strong consequences.

Lemma 9.76 *If C-meaningfulness and T-meaningfulness are both satisfied, then the order-invariance of one code extends to all the codes of a collection.*

For example, suppose that a quasi-upper convex T-class Γ^+ is right meaningful with respect to a meaningful $\langle LF \rangle$ -collection \mathcal{L}_c . If the initial code L is right order-invariant with respect to the family $\{\xi_{\lambda;1} \in \Gamma^+ \mid \lambda \in]0, 1[\}$, then, for any $\beta \in \mathbb{R}_{++}$, the code $L_{\alpha,\beta}$ in \mathcal{L}_c is right order-invariant with respect to the family $\{\xi_{\lambda;\beta} \in \Gamma^+ \mid \lambda \in]0, 1[\}$. Similar results hold for quasi-lower convex and quasi-relative translation T-classes and for T-families.

For a proof, see Falmagne and Doble (2015, p. 121).

Given a collection of codes, the only T-families or T-classes we are interested in are those which are meaningful with respect to that collection. So, it makes sense to link the meaningfulness of a T-family or a T-class with respect to a collection, and the corresponding order-invariance of the collection with respect to the T-family or the T-class. In view of Lemma 9.76, we can afford to require only the initial code's order-invariance. We introduce the following terminology.

Definition 9.77 Let \mathbf{Q} be a T-class which is either quasi-upper convex, quasi-lower convex, or a quasi-relative translation. The T-class \mathbf{Q} is a *legitimate right operator* for a collection \mathcal{W} if the following two conditions are satisfied:

- (i) \mathbf{Q} is right meaningful with respect to \mathcal{W} .
- (ii) The initial code of \mathcal{W} is right order-invariant with respect to the subclass of \mathbf{Q} which is:
 - (a) either of the form $\{\xi_{\lambda;\beta} \in \mathbf{Q} \mid \beta = 1, 0 < \lambda < 1\}$ in the quasi-upper convex or quasi-lower convex cases,
 - (b) or of the form $\{\eta_{\lambda;\beta} \in \mathbf{Q} \mid \beta = 1, \lambda \in \mathbb{R}_{++}\}$ in the quasi-relative translation case.

Similarly, a T-family \mathcal{F} is a *legitimate left (resp., right) operator* for a collection \mathcal{W} if \mathcal{F} is left (resp., right) meaningful with respect to \mathcal{W} , which has its initial code left (resp., right) order-invariant with respect to \mathcal{F} .

Another consequence, in addition to that given in Lemma 9.76, of combined C-meaningfulness and T-meaningfulness is that the relevant T-family or T-class is considerably restricted. This is shown in the next two lemmas.

Definition 9.78 A T-family \mathcal{F} on J is *1-point transitive* if there exists $\ell_0 \in J$ such that, for any $\lambda \in J$, we have $f(\ell_0) = \lambda$ for some $f \in \mathcal{F}$.

Lemma 9.79 Suppose that a 1-point transitive T-family \mathcal{F} on $[0, \infty[$ is a legitimate left operator for a meaningful $\langle LF \rangle$ -collection \mathcal{L}_c . Then \mathcal{F} is the similarity family; that is, $\mathcal{F} = \mathcal{S} = \{\sigma_\lambda \mid \sigma_\lambda : \ell \mapsto \lambda\ell, \lambda \in \mathbb{R}_{++}\}$. Accordingly, any code in \mathcal{L}_c has a multiplicative representation. More precisely, we must have for all $L_{\alpha,\beta} \in \mathcal{L}_c$ with $L_{\alpha,\beta} : [0, \infty[\times [0, \beta c] \rightarrow [0, \infty[$,

$$L_{\alpha,\beta}(\ell, v) = \alpha G\left(\left(\frac{\ell}{\alpha}\right)^\theta h\left(\frac{v}{\beta}\right)\right) \quad (\alpha, \beta, \theta \in \mathbb{R}_{++}), \quad (9.148)$$

with $h : [0, c] \rightarrow [0, \infty[$ and $G : [0, \infty[\rightarrow [0, \infty[$ two continuous, strictly monotonic functions which do not depend on α or β , with h decreasing and G increasing.

Remark 9.80 Similar results obviously hold for the collections relevant to Beer's Law (see Subsection 9.12.6 and Theorem 9.84) and to those relevant to the monomial laws in two variables (see Subsection 9.12.7 and Theorem 9.86), in the latter case for legitimate left and right operators.

The situation is slightly more complicated for T-classes. We give an example in the upper convex case.

Lemma 9.81 Let $\Gamma_{[0,c];u}^+$ be a quasi-upper convex T-class; thus, $\Gamma_{[0,c];u}^+$ contains all the transformations of the form

$$\begin{aligned} \xi_{\lambda;\beta}(v) &= u^{-1}(\lambda u(v) + (1 - \lambda)u(\beta c)), \\ (0 < \lambda < 1, \beta &\in \mathbb{R}_{++}, 0 \leq v \leq c), \end{aligned} \quad (9.149)$$

with $u : [0, \infty[\rightarrow [0, \infty[$ strictly increasing and continuous and $c \in \mathbb{R}_{++}$, a constant. Suppose that $\Gamma_{[0,c];u}^+$ is a legitimate right operator for a meaningful $\langle LF \rangle$ -collection \mathcal{L}_c . Then $u(v) = av^\theta + b$ for some constants a, θ and b in \mathbb{R} , with $a\theta \in \mathbb{R}_{++}$. Thus, $u(0) = b$ and the transformation $\xi_{\lambda;\beta}(v)$ has the form

$$\xi_{\lambda;\beta}(v) = (\lambda v^\theta + (1 - \lambda)(\beta c)^\theta)^{1/\theta}. \quad (9.150)$$

Note in passing that the solution $u(v) = a \ln v + b$ of the functional equation (9.149) is declined because the function u is assumed to be defined at the point 0.

9.12.5 The Lorentz–FitzGerald Contraction

Recall the LF-Contraction encountered earlier in Part II of this chapter:

$$[\text{LF}] \quad L(\ell, v) = \ell \sqrt{1 - \left(\frac{v}{c}\right)^2}, \quad (0 \leq \ell < \infty, 0 \leq v \leq c), \quad (9.151)$$

where ℓ is the length of a rod measured by an observer at rest with respect to the rod and $L(\ell, v)$ denotes the length of the rod measured by an observer moving at the speed v with respect to the rod.

We now give the representation theorem dealing with the LF-Contraction. In particular, we derive a mathematical form of the LF-Contraction, up to two numerical parameters, from four invariance axioms. The first three axioms involve C-meaningfulness and T-meaningfulness, and the last one specifies an initial condition. For a proof, see Falmagne and Doble (2015, pp. 125–127).

Theorem 9.82 *Let $\mathcal{L}_c = \{L_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ be a $\langle LF \rangle$ -collection of codes, with $L_{\alpha,\beta} : [0, \infty[\times [0, \beta c] \rightarrow [0, \infty[$ and $c \in \mathbb{R}_{++}$, a constant. Consider the following four axioms:*

- [LF1] *The collection \mathcal{L}_c is ST-meaningful (see Definition 9.28).*
- [LF2] *There is a 1-point transitive T-family \mathcal{F} which is a legitimate left operator for \mathcal{L}_c .*
- [LF3] *For some function $u : [0, \infty[\rightarrow [0, \infty[$, strictly increasing and continuous, the quasi-upper convex T-class $\Gamma_{[0,c];u}^+ = \cup_{\beta \in \mathbb{R}_{++}} u\mathcal{C}_{[0,\beta c]}^+u^{-1}$ is a legitimate right operator for \mathcal{L}_c .*
- [LF4] *$L(\ell, 0) = \ell$ for all $\ell \in [0, \infty[$.*

Axioms [LF1]–[LF4] hold for the collection \mathcal{L}_c if and only if there exist two positive constants ξ and ψ such that, for all codes $L_{\alpha,\beta} \in \mathcal{L}_c$,

$$L_{\alpha,\beta}(\ell, v) = \ell \left(1 - \left(\frac{v}{\beta c}\right)^\psi\right)^\xi, \quad (0 \leq \ell < \infty, 0 \leq v \leq \beta c). \quad (9.152)$$

The standard form of the LF-contraction is obtained when $\psi = 2 = \frac{1}{\xi}$; the initial code L (that is, $\alpha = \beta = 1$ in (9.152)) then takes the form of Equation (9.151).

9.12.6 Beer's Law

We recall that this law applies in a class of situations where an incident radiation traverses some absorbing medium, so that only a fraction of the radiation goes through. In our notation, the expression of the law is

$$I(x, y) = xe^{-\frac{y}{c}}, \quad (9.153)$$

in which x denotes the intensity of the incident light, y is the concentration of the absorbing medium, c is a reference level, and $I(x, y)$ is the intensity of the transmitted radiation. For convenience, we recall the basic facts about the corresponding collection of codes.

Definition 9.83 For all $\alpha, \beta \in \mathbb{R}_{++}$, let $I_{\alpha,\beta} : [0, \infty[\times [0, \infty[\rightarrow [0, \infty[$ be a numerical code, which is strictly increasing in its first variable and strictly decreasing in its second variable. Writing $I = I_{1,1}$ for the initial code, we suppose that the scales for $I_{\alpha,\beta}(x, y)$ and x are identical, and that the unit for each is specified by α . The set $\mathcal{I} = \{I_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ of all such codes is referred to as a $\langle B \rangle$ -collection.

What follows is a representation theorem dealing with Beer's Law. For a proof, see Falmagne and Doble (2015, pp. 129–130). Note that axioms [BL1]–[BL4] for this representation below resemble [LF1]–[LF4] for the LF-contraction representation. The only essential difference lies in Axiom [BL3], which captures the exponential operator in (9.153) and the special role of the constant c . It is not surprising that the representations would be similar, as the forms of both Beer's law (Equation (9.153)) and the LF-contraction (Equation (9.151)) involve the reduction of an input quantity by a multiplicative operator.

Theorem 9.84 Let $\mathcal{I} = \{I_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ be a $\langle B \rangle$ -collection of codes, with $I_{\alpha,\beta} : [0, \infty[\times [0, \infty[\rightarrow [0, \infty[$ and $c \in \mathbb{R}_{++}$, a constant. Consider the following four axioms:

- [BL1] The collection \mathcal{I} is ST-meaningful (see Definition 9.28).
- [BL2] There is a 1-point transitive T -family \mathcal{F} which is a legitimate left operator for \mathcal{I} .
- [BL3] For some strictly increasing and continuous $u : [0, \infty[\rightarrow [0, \infty[$, the quasi- c -relative translation T -class $\mathbf{T}_{c;u} = \cup_{\beta \in \mathbb{R}_{++}} uT_{\beta c}u^{-1}$ is a legitimate right operator for \mathcal{I} .
- [BL4] $I(x, c) = \frac{x}{e}$ for all $x \in [0, \infty[$, where I is the initial code of the collection \mathcal{I} , and the constant c is as in [BL3].

Axioms [BL1]–[BL4] hold for the collection \mathcal{I} if and only if there exist two positive constants θ and δ such that, for all codes $I_{\alpha,\beta} \in \mathcal{I}$,

$$I_{\alpha,\beta}(x, y) = \frac{x\theta^{-\left(\frac{y}{\beta c}\right)^{\delta}+1}}{e}, \quad (x, y \in [0, \infty[). \quad (9.154)$$

This equation becomes Beer's law (Equation (9.153)) if we set $\delta = \alpha = \beta = 1$ and $\theta = e$.

9.12.7 The Monomial Laws

Monomial laws are used to describe a wide variety of scientific phenomena. Although for the monomial laws the output scale may be different from the two

input scales, such laws can be axiomatized along the lines of the LF-contraction (Theorem 9.82) and Beer's law (Theorem 9.84).

Definition 9.85 For all $\alpha, \beta \in \mathbb{R}_{++}$, let $Q_{\alpha,\beta} :]0, \infty[\times]0, \infty[\rightarrow]0, \infty[$ be a numerical code, with α and β specifying the units of the two input variables. We also define $Q = Q_{1,1}$ as the initial code. The collection of functions $\mathcal{Q} = \{Q_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ is called a $\langle M \rangle$ -collection.

We omit the proof of the theorem below (see Falmagne & Doble, 2015, pp. 131–132).

Theorem 9.86 Let $\mathcal{Q} = \{Q_{\alpha,\beta} \mid \alpha, \beta \in \mathbb{R}_{++}\}$ be a $\langle M \rangle$ -collection of codes, with $Q_{\alpha,\beta} :]0, \infty[\times]0, \infty[\rightarrow]0, \infty[$. Consider the following four axioms:

- [ML1] The collection \mathcal{Q} is (δ_1, δ_2) -meaningful (see Definition 9.27).
- [ML2] There is a 1-point transitive T-family \mathcal{F} which is a legitimate left operator for \mathcal{Q} .
- [ML3] There is a 1-point transitive T-family \mathcal{G} which is a legitimate right operator for \mathcal{Q} .
- [ML4] The function $x \mapsto Q(x, 1)$ is homogeneous of some degree $\mu \neq 0$, that is, $Q(\lambda x, 1) = \lambda^\mu Q(x, 1)$ for all $x, \lambda \in]0, \infty[$.

Axioms [ML1]–[ML4] hold for the collection \mathcal{Q} if and only if there exist constants $\zeta, \gamma \neq 0$ such that, for all codes $Q_{\alpha,\beta} \in \mathcal{Q}$,

$$Q_{\alpha,\beta}(x, y) = \alpha^{\delta_1 - \mu} \beta^{\delta_2 - \gamma} \zeta x^\mu y^\gamma \quad (x, y \in]0, \infty[), \quad (9.155)$$

with the signs of all constants consistent with the monotonicity of the codes in the two variables.

9.12.8 A Counterexample: Van der Waals' Equation

Not all scientific laws can be axiomatized by the methods of this section, as some scientific laws are not order-invariant with respect to non-trivial transformations of one of their input variables. A case in point is van der Waals' equation, which was introduced in Example 9.18(d):

$$T(p, v) = K \left(p + \frac{a}{v^2} \right) (v - b), \quad (p \in]0, \infty[, v \in [b, \infty[), \quad (9.156)$$

in which p is the pressure of a fluid, v is the volume of the container, T is the temperature, a and b are parameters, and K is the reciprocal of the Boltzmann constant.

We suppose that $T : \mathbb{R}_{++} \times [b, \infty[\rightarrow \mathbb{R}_{++}$ is a code in a collection \mathcal{T} . So, implied in (9.156) are three indices specifying the three ratio scales used to measure the pressure p , the volume v and the absolute temperature T . The quantities K , a , and b are functions of these indices.

Notice that, with an appropriate choice of K , a , and b , van der Waals' equation in the form of Equation (9.156) is meaningful in the sense of Definition 9.27; that is, we have

$$\frac{1}{\gamma} T_{\alpha,\beta,\gamma}(\alpha p, \beta v) = \frac{1}{\delta} T_{\theta,\eta,\delta}(\theta p, \eta v). \quad (9.157)$$

To see this, take

$$K = \frac{\gamma}{\alpha\beta}, \quad a = \beta^2\alpha, \quad b = \beta, \quad (\alpha, \beta, \gamma, \theta, \eta, \delta \in \mathbb{R}_{++}).$$

Then, with appropriate change of notation, (9.156) becomes

$$\frac{1}{\gamma} T_{\alpha,\beta,\gamma}(\alpha p, \beta v) = \frac{1}{\gamma} \left(\frac{\gamma}{\alpha\beta} \right) \left(\alpha p + \frac{\beta^2\alpha}{(\beta v)^2} \right) (\beta v - \beta) = \left(p + \frac{1}{v^2} \right) (v - 1),$$

which does not depend upon α , β or γ . So, (9.157) holds, and the corresponding collection of codes is meaningful. (Note that the r.h.s. of the above formula is equal to $T_{1,1,1}$.)

However, (9.156) is neither left nor right order-invariant with respect to non-trivial (non-identity) transformations. Indeed, suppose first that T is left order-invariant. Then there exists a transformation $f : [0, \infty[\rightarrow [0, \infty[$ (with f fixed for this argument) and, by Theorem 9.64, a strictly increasing continuous function $\phi = \phi_f$ such that

$$\begin{aligned} \phi \left(K \left(p + \frac{a}{v^2} \right) (v - b) \right) &= K \left(f(p) + \frac{a}{v^2} \right) (v - b), \\ (p \in [0, \infty[, v \in [b, \infty[). \end{aligned} \quad (9.158)$$

Because the r.h.s. is differentiable in v , so is the l.h.s. Hence, the function ϕ is differentiable, and thus f is also differentiable. Taking derivatives on both sides with respect to p , we obtain after simplification, for $v > b$,

$$\phi' \left(K \left(p + \frac{a}{v^2} \right) (v - b) \right) = f'(p).$$

Since the r.h.s. does not vary with v , ϕ' must be a constant function. Thus, both f and ϕ are linear or affine functions. Considering (9.158), it is easily verified that the only possibility is that both ϕ and f are identity functions. In other words, T cannot be left order-invariant with respect to non-trivial transformations.

The argument pertaining to right order-invariance is similar. If T is right order-invariant, there exist two continuous, strictly increasing functions $\xi : [0, b] \rightarrow [0, b]$ and ϕ such that

$$\phi \left(K \left(p + \frac{a}{v^2} \right) (v - b) \right) = K \left(p + \frac{a}{\xi(v)^2} \right) (\xi(v) - b). \quad (9.159)$$

The r.h.s. of (9.159) is affine in p , so the l.h.s. must be also. We obtain, for some constants α and μ ,

$$\alpha \left(K \left(p + \frac{a}{v^2} \right) (v - b) \right) + \mu = K \left(p + \frac{a}{\xi(v)^2} \right) (\xi(v) - b). \quad (9.160)$$

Taking the derivatives with respect to p on both sides yields

$$\alpha K(v - b) = K(\xi(v) - b).$$

So, ξ is linear. Replacing ξ in (9.160) by its linear expression, it follows that ξ must be the identity function on $[0, b]$. We conclude that van der Waals' equation, in the form (9.156), is neither left nor right order-invariant with respect to non-trivial transformations. However, we do not know whether it is order-invariant with respect to a family of transformations jointly affecting its two input variables.

This concludes our discussion of meaningfulness as an invariance principle applied to families of real-valued functions and the axiomatization of scientific or mathematical laws given by such functions. We now turn to a different approach to meaningfulness, a set-theoretic view of invariance in science and mathematics that aims to overcome the shortcomings of Klein's Erlanger program (discussed in Section 9.4) and the representational theory of measurement (discussed in Section 9.3).

Part III: Meaningfulness and Scientific and Geometric Content

9.13 Principles for Scientific and Geometric Content

The Erlanger program (see Klein, 1872, and Section 9.4 of this chapter) has two great weaknesses: (1) the inability to effectively deal with situations that had the identity as its only symmetry, and (2) the lack of an argument justifying identifying geometrical content with invariance under symmetry groups. The representational theory has the same two weaknesses.

Narens (1988, 2002a, 2015) developed a radically different approach to meaningfulness. Instead of basing meaningfulness on some invariance concept, he focused on the Erlanger's use of invariants for defining a geometry's content or subject matter. He writes:

Fundamental to the theories [of measurement] is that certain relations and concepts belong to a fragment of science and others do not. Those that belong are called "meaningful" (with respect to the fragment) and those that do not are called "meaningless" (with respect to the fragment). A theory of meaningfulness consists of giving necessary conditions for meaningfulness. Because theories of meaningfulness are metascientific and philosophical, different theories arise naturally out of various epistemological and metaphysical positions typically taken about science. Several formal theories of meaningfulness are presented ... [that are] discussed in Narens (2002a). The framework and theories presented in this and later chapters are simplified descriptions of a few of those. The theories are based on precise descriptions of "scientific topic" and "meaningfulness." They are designed to satisfy the following five intuitive principles.

Principle 1. The domain of the scientific topic is a qualitative set X .

Principle 2. The scientific topic is determined by a structure of primitives $(X, Q_j)_{j \in J}$, where each Q_j is a higher-order relation based on X .

Principle 3. The structure of primitives and each of its primitives belong to the scientific topic.

Principle 4. The scientific topic is closed under “scientific definition,” that is, if b_1, \dots, b_n belong to the topic and b is defined “scientifically” in terms of b_1, \dots, b_n , then b belongs to the scientific topic.

Principle 5. A portion of pure mathematics can be used in scientific definitions.

As an example of Principles 1 to 3, consider the geometry of space-time in relativistic physics. This consists of a domain X of space-time points with the relativistic distance function ρ on X . X is also the domain for many other qualitative geometries, for example, X with an appropriately defined 4-dimensional Euclidean metric, δ_4 , on it is a 4-dimensional Euclidean geometry, and X with an appropriately defined 2-dimensional Euclidean distance function, δ_2 , on it is 2-dimensional Euclidean geometry. Each of these geometries is viewed as a different scientific topic. Many 4-dimensional Euclidean concepts about relativistic space-time are considered to be outside of space-time relativity theory; they are considered to be “nonrelativistic” – or to use the language of this chapter – they are meaningless in the sense that they have no “relativistic interpretation.” Many thought experiments in relativistic physics employ in informal ways this use of “meaningless.” The meaningful/meaningless distinction also appears as a major ingredient in other methods of inference used in physics, particularly in hypothesis formation and the design of experiments.

Principle 5 reflects a common practice in mathematical science of utilizing a portion of pure mathematics in the formulations of scientific concepts and the derivations of scientific results. Because of the non-empirical, non-qualitative nature of pure mathematics, the combination of Principles 3 to 5 allow relations and concepts to belong to the scientific topic that are neither qualitative nor empirical. Thus a concept of “meaningfulness,” based in part on Principles 3 to 5, should not be identified with either “empiricalness” or “qualitativeness.” In the literature, identifications of meaningfulness concepts with qualitativeness or empiricalness have caused confusion. . . .

The portion of mathematics allowable in Principle 5 depends, in part, on the investigator’s philosophies of science and mathematics. For example, some may believe that entities of pure mathematics whose existence depend on the Axiom of Choice – a highly nonconstructive and infinitistic principle of set theory – should not be allowed in determining the entities that belong to the topic of science under consideration, whereas the use of real numbers and various operations on them are allowed. Others may believe that all of ordinary mathematics may be used in describing what belongs to the topic. In principle, the portion of mathematics referred to in Principle 5 may range from none to all of ordinary mathematics.

(Narens, 2007, pp. 115–116.)

Narens (2002a) carried out a program based on Principles 1 to 5 and related it to the Erlanger program. In order to accomplish this, he needed a formal system that could describe higher-order qualitative relations based on a non-empty domain X , a qualitative measuring structure having as primitives such qualitative relations, relations and concepts of all of pure mathematics (considered to be non-qualitative),

and higher-order relations based on domains consisting of higher-order qualitative relations and higher-order relations of pure mathematics. (An example of the latter is an isomorphism of the qualitative measuring structure onto a numerically based measuring structure.) Also needed was a method of defining in such a complicated higher-order system what “invariants” are so that the resulting meaningfulness concept could be related to the Erlanger program as well as the various meaningfulness concepts of the representational theory. Narens did this by extending the axiomatic system ZFC, which is the Zermelo–Fraenkel set theory with the axiom of choice, to include a set of non-sets as the domain of the qualitative measuring structure, and a new predicate $M(x)$ (read as: “ x is meaningful”) and providing additional axioms for these new concepts.

The set theory used for the theories of meaningfulness developed in this section is called **ZFCAM**. It is a structure of the form,

$$\mathfrak{V} = \langle V, \in, A, \emptyset, M \rangle,$$

where V is a collection of *entities*, \in is the set-theoretic membership relation, A is a special set of entities, \emptyset is the empty set, and M is a subcollection of V . Each entity of V is either an atom or a set. The object x is said to be a *set* if and only if x is the empty set \emptyset , or there exists y such that $y \in x$. The empty set has no elements. It is assumed that A is different from \emptyset . Each element of A is called an *atom*, and A is called the *set of atoms*. Atoms are non-sets, that is, the empty set is not an atom and none of the atoms is a set. ZFCAM’s axioms consist of the axioms for ZFC with the definition of the empty set \emptyset , and the axiom of foundation is modified to accommodate atoms as follows:

- (i) $\emptyset \neq A$ and for each x , either $x \notin \emptyset$ or $x \in A$; and
- (ii) for each x , if x is a set and $x \neq \emptyset$, then there exists y such that $y \in x$ and either $y \in A$ or $y \cap x = \emptyset$.

The collection M is the interpretation of M . This is the general theory for ZFCAM. Specific theories involving it will be given later, which have additional axioms involving M . They will be theories of meaningfulness designed to meet the Principles 1 to 5 above.

ZFC only has sets as its entities. In the foundations of mathematics, it is assumed to comprise all of pure mathematics. That is, the theory of sets belongs to pure mathematics and all pure mathematics have isomorphic counterparts in ZFC. The collection V will be so defined that it contains all the sets of ZFC, that is $\mathfrak{V} = \langle V, \in, A, \emptyset, M \rangle$ contains all of pure mathematics.

For each non-negative integer n , define the sets V_n and P_n through induction as follows:

- For $n = 0$, $V_0 = A$ and $P_0 = \emptyset$.
- If V_n has been defined, then

$$V_{n+1} = \wp(V_n) \cup V_n \text{ and } P_{n+1} = \wp(P_n) \cup P_n,$$

where \wp is the power set function.

- By definition,

$$V_\infty = \bigcup_{k=0}^{\infty} V_k \text{ and } P_\infty = \bigcup_{k=0}^{\infty} P_k.$$

In set theory notation, $V_\infty = V_\omega$, where ω is the first infinite ordinal and similarly for P_∞ . Note that all the elements in P_∞ are pure sets.

If A is infinite, then V_∞ is a very large and rich set. However, in order to obtain a very general meaningfulness concept, a still larger collection is needed. This can be done through transfinite induction on ordinals in a manner similar to the finite induction for V_∞ and P_∞ above, to arrive at the necessary collections V and P , that is,

$$V_{\omega+1} = \wp(V_\omega) \cup V_\omega \text{ and } P_{\omega+1} = \wp(P_\omega) \cup P_\omega, \text{ and so on through all ordinals.}$$

It is not difficult to show that P consists of all the sets in ZFC, which here is interpreted as all of pure mathematics.

The collection V is called the *set of scientific entities* and P is called the *set of purely mathematical entities*. It is easy to show $P \subset V$. Thus \mathfrak{V} satisfies Principle 5 with the portion of pure mathematics being available for scientific definitions consisting of all of pure mathematics.

Suppose X is a non-empty set. Then R is said to be a *first-order* relation on X if $R \neq \emptyset$ and R is an n -ary relation on X for some non-negative integer n . Other kinds of relations built up out of first-order relations on X are called *higher-order* relations on X . These include second-order relations on X , for example, a non-empty n -ary relation on a non-empty subset of first-order relations, third-order relations on X , for example, a non-empty n -ary relation on a non-empty subset of second-order relations on X , etc. Higher-order relations include first-order relations. Those higher-order relations that are not first-order are considered more abstract than first-order relations. Levels of abstraction among higher-order relations can be formulated. However, for the purposes of this chapter, only the distinction between first- and higher-order is needed.

Note that by the above usage, all first-order relations are higher-order, but some higher-order relations are not first-order.

In the foundations of science, the fragment of science under consideration is often characterized as a structure of the form $\langle A, R_j \rangle_{j \in J}$, where for each $j \in J$, R_j is a higher-order relation on A . In most characterizations, the R_j are first-order and observable. Higher-order relations that are not first-order, if they appear at all, usually appear in some axiom about the fragment of science under consideration. For example, in the characterization of a continuous extensive structure, $\langle A, \precsim, \oplus \rangle$, the relations \precsim and \oplus are first-order relations. However, higher-order relations that are not first-order appear as part of the axiom of continuous domain.

The collection V contains all higher-order relations based on A . This is because it is well-known that the concept of an “ n -ary relation” among entities can be formulated in terms of the primitives of ZFCAM.

ZFC is formulated in the first-order language $\mathcal{L}(\in)$. ZFCAM uses the broader language $\mathcal{L}(\in, \mathbf{A}, \emptyset, \mathbf{M})$. Note that this broader first-order language includes as a sublanguage the first-order language for ZFC.

The first meaningfulness theory considered is axiom system TM. It is the Erlanger program's idea of "geometric content" formulated in ZFCAM:

TM: There exists a transformation group H on A such that for each entity e and each h in H ,

$$h(e) = e \iff e \in \mathbf{M}.$$

As a theory of scientific content, TM has the same two serious weaknesses as Klein's Erlanger Program:

- (1) justifying invariance under symmetries as the criterion for scientific content, and
- (2) requiring that the scientific content for each situation having the identity transformation as its only symmetry contains all higher-order relations on the qualitative domain, A .

For geometry, Klein (1872) rejected as true geometries metric spaces that resulted from the differentiation methods proposed by Riemann that had the identity as its only symmetry. However, such a view became untenable with the advent of Einstein's theory of general relativity (Feynman, Leighton, & Sands, 1963, vol. 2), because it rejected physical space-time as a geometry. There were various attempts by geometers to generalize the Erlanger program to take into account some methods of differential geometry that produced geometries with single symmetries, e.g., investigating invariance under various kinds of local symmetries, but these accounts encountered severe difficulties as generalizations of the Erlanger program and never had much of an impact on geometrical thinking. Measurement theorists, particularly Duncan Luce (1978), recognized (2) as a serious issue for representational theories of meaningfulness; (1) is obviously a serious philosophical issue. It never became a serious mathematical one, although occasionally vague rationale for its use was suggested. It became accepted as a principle because it – and the rest of the Erlanger program – led to good insights and productive geometric methods. The use of the following transfer principle described in Veblen and Young (1946) is an example.

At each step we have helped ourselves forward by transferring the results of one geometry to another, combining these with easily obtained theorems of the second geometry, and thus extending our knowledge of both. This is one of the characteristic methods of modern geometry. It was perhaps first used with a clear understanding by O. Hesse [*Gesammelte Werke*, p. 531], and was formulated as a definite geometrical principle (*Übertragungsprinzip*) by F. Klein (1872).

This principle of transference or of carrying over the results of one geometry to another may be stated as follows:

Given a set of elements $[e]$ and a group G of permutations of these elements, and a set of theorems $[T]$ in which the relations are left invariant by G . Let $[e']$ be another set of elements, and G' a group of permutations of $[e']$. If there is a one-to-one reciprocal correspondence between $[e]$ and $[e']$ in which G is simply isomorphic with G' , the set of theorems $[T]$ determines by a mere change of terminology a set of theorems $[T']$ which state relations among elements e' which are left invariant by G' .

This principle becomes effective when the method by which $[e]$ and G are defined is such as to make it easy to derive theorems which are not so easily seen for $[e']$ and G' . This has been abundantly illustrated in the present chapter ...

DEFINITION. Given a set of elements $[e]$ and a group G of permutations of $[e]$, the set of theorems $[T]$ which state relations among the elements of $[e]$ which are left invariant by G and are not left invariant by any group of transformations containing G is called a *generalized geometry* or a *branch of mathematics*.

This is, of course, a generalization of the definition of a geometry by Klein (1872, pp. 284–285).

(Veblen & Young, 1946).

Narens (2002a) generalizes the Erlanger program by finding conjunctions of propositions about \mathfrak{V} that are logically equivalent to $T\mathcal{M}$, and then deleting some conjuncts to get more general systems of meaningfulness. The following lemmas are consequences of axiom system $T\mathcal{M}$. Most are straightforward consequences. Their proofs can be found in Narens (2002a).

Lemma 9.87 $T\mathcal{M}$ implies $M(A)$ and $M(\emptyset)$.

Definition 9.88 The MP (*meaningful pure sets*) axiom is said to hold if for each pure set x in V , $M(x)$ is true.

Lemma 9.89 $T\mathcal{M}$ implies MP.

Definition 9.90 Assume $T\mathcal{M}$. Then axiom MI (*meaningful inheritability*) holds if for each set x in V , x is meaningful if all elements of x are meaningful.

Lemma 9.91 $T\mathcal{M}$ implies MI.

Definition 9.92 Axiom AL (*atomic legacy*) is said to hold if for each a in A , if $M(\{a\})$, then $M(a)$.

Lemma 9.93 $T\mathcal{M}$ implies AL.

Definition 9.94 Assume $T\mathcal{M}$. Axiom MC' (*meaningful comprehension'*) is said to hold if for all formulas $\Phi(x, u_1, \dots, u_n)$ of the first-order language $\mathcal{L}(\in, A, \emptyset)$ for $\langle V, \in, A, \emptyset \rangle$ and all entities e, e_1, \dots, e_n in V :

if e_1, \dots, e_n are meaningful and $e = \{x | \Phi(x, e_1, \dots, e_n)\}$, then $M(e)$.

Note that in Definition (9.94), Φ is a formula of $\mathcal{L}(\in, A, \emptyset)$ and therefore does not contain the predicate M . The axiom of meaningful comprehension' allows for

the definition of new meaningful entities in terms of already known ones. It is a principle very similar to the axiom of comprehension of ZFC.

A proof of the following lemma follows from theorem 4.2.5 of Narens (2002a).

Lemma 9.95 *TM implies MC'.*

Theorem 9.96 *Axiom TM holds if and only if the six following conditions hold: M(A), M(\emptyset), MP, MI, AL, and MC'.*

For a proof, see Narens (2002a, theorem 4.2.10).

Theorem 9.96 suggests a method for generalizing the Erlanger program concept of “meaningfulness”: drop or modify one or more of the conjuncts on the r.h.s. of “if and only if” in the statement of the theorem. Narens (2002a) considered many options for this, each producing an acceptable and interesting theory of meaningfulness. This chapter follows one of these – axiom system D' – that drops MI. Axiom MI is not really a definability or constructibility principle, because it implies that a set that exists only through an application of the axiom of choice that has only meaningful entities is itself meaningful.

Definition 9.97 Axiom system D' consists of MP, AL, and MC'.

It easily follows that axiom system D' implies M(A) and M(\emptyset).

Theorem 9.96 shows that axiom system D' is implied by TM. Theorem 4.3.7 of Narens (2002a) shows that D' does not imply TM. Thus D' is a proper generalization of TM.

Narens (2002a) has the following comments about D':

Throughout this book, axiom system D' and its extensions will often be referred to as *definitional systems of meaningfulness*. Strictly speaking, axiom system D' is not a form of *logical* definability, because it assumes the axiom of Meaningful Pure Sets. However, it appears to me to capture a form of “definability” that is commonly used in science, if it is accepted that pure sets corresponds to pure mathematics. For example, in science qualitative relations are often “defined” in terms of other qualitative objects and real numbers. Axiom system D' is intended to capture this and similar practices of *scientific* definability. At various places in the book, the kind of “definability” inherent in axiom system D' is called “set-theoretic definability,” in part, to distinguish it from “logical definability.”

(Narens, 2002a, p. 143.)

Axiom TM provides necessary and sufficient conditions for meaningfulness; DM' provides only necessary conditions.

Because knowing the non-meaningfulness of specific entities is often important in applications, it is useful to have criteria for non-meaningfulness under DM'. One method for accomplishing this is through Lemma 9.95. Consider the situation where there is a structure of primitives $\mathfrak{Q} = \langle A, R_j \rangle_{j \in J}$ with J being a pure set that determines a qualitative situation. Assume the relation R belongs to the scientific content based on \mathfrak{Q} , where this is taken to mean that each R_j is meaningful and axiom system D' holds. Let H be the symmetry group for \mathfrak{Q} . Then it follows from TM implies DM' (Theorem 9.96) that if an entity is not invariant under H , then it

is meaningless. The following quotations of Narens (1988) and Narens (2002b) show how this result provides a foundation for deriving propositions for dimensional analysis and related methods based on invariance. (The quotations have been altered to refer to theorems and definitions of this section.)

Suppose in a particular setting we are interested in finding the functional relationship of the qualitative variables, x , y , and z . We believe that the primitive relations (which are known) completely characterize the current situation. Furthermore, our understanding (or insight) about the situation tells us that x must be a function of y and z . (This is the typical case for an application of dimensional analysis in physics.) This unknown function – which we will call “the desired function” – must be determined by the primitives and the qualitative variables x , y , and z . Therefore, it should somehow be “definable” from these relations and variables. Even though the exact nature of the definability condition is not known, (it can be argued that) it must be at the same level or weaker than the enormously powerful methods of definability encompassed by Axiom System TM. Thus [by knowing that it is at the same level or weaker than TM] we know that any function relating the variable x to the variables y and z that is *not* invariant under the symmetries of the primitives cannot be the desired function.

In many situations, this knowledge of knowing that functions not invariant under the symmetries of the primitives cannot be the desired function can be used to effectively find or narrow down the possibilities for the desired function.

(Narens, 1988, pp. 70–71.)

Narens (2002b) expanded on this theme as follows:

Scientific inquiry is a complicated issue with many overlapping parts. I believe meaningfulness belongs primarily to the theoretical part of scientific inquiry. Because of the overlap of the theoretical part of a science with its experimental and applied parts, meaningfulness often has important ramifications in the experimental and applied parts. Meaningfulness is essentially a theoretical position about scientific content and its role in (theoretical) inference. For example, consider the case where by extra-scientific means (e.g., intuition, experience, etc.) a scientist is led to believe that a function $z = F(x, y)$ that he needs to describe from a subset of $A \times A$ into A is completely determined by the observable, first-order relations R_1, \dots, R_n on A . Then it is reasonable for the scientist to proceed under the hypothesis that F belongs to the scientific content of $\mathfrak{X} = \langle A, R_1, \dots, R_n \rangle$, which for this discussion may be taken as the set of meaningful entities determined by Axiom System D'. Thus the scientist assumes F has a scientific definition in terms of \mathfrak{X} and its primitives. By Lemma 9.95, F is invariant under the of \mathfrak{X} . Suppose the scientist knows enough properties about \mathfrak{X} and has the mathematical skill to determine the symmetry group G of \mathfrak{X} . Then methods of analyses involving symmetries may be employed to provide information helpful in characterizing F . There are several methods in the literature for accomplishing this.

Note that in the above process, scientific definability is used to *justify* F belonging to the appropriate topic, invariance is used as a mathematical technique to find helpful information for characterizing F , and that these two uses are connected by a theorem of mathematical logic. Also note that the scientist's belief that F belonged to the topic generated by \mathfrak{X} is extra-scientific. Therefore, the deductions based on information obtained through the above process should either be checked by experiment or be derived from accepted scientific theory and facts; i.e., they

should be treated as scientific hypotheses that need corroboration. Thus, for the purposes of science, the above process is a method of generating hypotheses and not facts: If the scientist's extra-scientific beliefs are correct, then the generated hypotheses will be facts; however, the scientist has no *scientific guarantee* that his beliefs are correct.

(Narens, 2002b, pp. 764–765.)

9.13.1 Intrinsicness

Intrinsicness is a stronger form of invariance than symmetry invariance. It was developed in Narens (2002a) as a concept (i) to capture a form of lawfulness, and (ii) for using D' as a meaningfulness concept for situations where the underlying qualitative structure has only a single symmetry. This subsection considers only a special case of (ii). For a general account of intrinsicness and its many applications, the reader is referred to chapter 5 of Narens (2002a).

Convention 3 Throughout the remainder of this section, the following assumptions are made:

- $\mathfrak{X} = \langle A, R_j \rangle_{j \in J}$ is a qualitative structure, with a pure set J .
- H is the symmetry group for \mathfrak{X} .
- Axiom system TM holds with

$$M(x) \text{ holding} \iff f(x) = x \text{ for all } f \text{ in } H \text{ and all entities } x \text{ in } V.$$

- $\mathcal{F}_{\mathfrak{X}}$ is a set of isomorphic structures with domain A and such that $\mathfrak{X} \in \mathcal{F}$.

The set $\mathcal{F}_{\mathfrak{X}}$ is called a *meaningful field of structures (based on \mathfrak{X})*.

Let $\mathcal{F}_{\mathfrak{X}}$ be a meaningful field of structures. Intuitively, the structure \mathfrak{X} is taken as a perspective that accounts for the scientific topic M under consideration. The accounting is done in terms of pure mathematics and a language appropriate to \mathfrak{X} . There are other perspectives for M that also validly account for the same scientific topic. The meaningful field $\mathcal{F}_{\mathfrak{X}}$ is a particular kind of set of such perspectives. What makes it particular is that all elements of $\mathcal{F}_{\mathfrak{X}}$ are isomorphic²⁵ to \mathfrak{X} . This allows for a common language, $L = L(\epsilon, A, \emptyset, R_j)_{j \in J}$, to be used for describing each structure in $\mathcal{F}_{\mathfrak{X}}$. The idea of intrinsicness is that some scientific entities have the same description in terms of L no matter which structure in $\mathcal{F}_{\mathfrak{X}}$ is being described by L . Such entities are called “intrinsic.” Intuitively, intrinsicness has an extra kind of invariance that is not necessarily captured by the symmetries of \mathfrak{X} . Another way of viewing this is that perspectives in \mathcal{F} are contexts, and intrinsic relations are meaningful relations that are invariant under change of context.

Definition 9.98 (The language $L_{\mathfrak{X}}$) By definition, $L_{\mathfrak{X}} = L(\epsilon, A, \emptyset, R_j)_{j \in J}$ is the extension of the first-order language $L(\epsilon, A, \emptyset)$ with the additional individual constant symbols \mathfrak{X} and R_j for $j \in J$.

²⁵ The more general concept of “intrinsicness” of Narens (2002a) relies on the structures in the meaningful field having a common language. They need not be isomorphic.

Convention 4 The individual constant symbols \mathbf{R}_j , for $j \in J$, and \mathfrak{X} of $\mathbf{L}_{\mathfrak{X}}$ are interpreted for elements \mathfrak{X}^* of \mathcal{F} as follows:

- If $\mathfrak{X}^* = \mathfrak{X}$, then \mathbf{R}_j is interpreted as R_j and \mathfrak{X} as \mathfrak{X} .
- If $\mathfrak{X}^* = \langle A, R_j^* \rangle_{j \in J}$, then \mathbf{R}_j is interpreted as R_j^* and \mathfrak{X} as \mathfrak{X}^* .

Let $\mathcal{F}_{\mathfrak{X}}$ be a field of meaningful structures and $R \in V$. Then R is said to be $\mathcal{F}_{\mathfrak{X}}$ -*intrinsic* if and only if there exists a formula $\Phi(x, x_1, \dots, x_m)$ of $\mathbf{L}_{\mathfrak{X}}$ and pure sets b_1, \dots, b_m in P such that the following is a true statement about each structure in \mathcal{F} :

$$\Phi(R, b_1, \dots, b_m) \text{ and for all } x, \text{ if } \Phi(x, b_1, \dots, b_m) \text{ then } x = R.$$

Theorem 9.99 *Let $\mathcal{F}_{\mathfrak{X}}$ be a field of structures. Define the new meaningfulness predicate M' for \mathfrak{X} by:*

$$M'(x) \iff M(x) \text{ and } x \text{ is } \mathcal{F}_{\mathfrak{X}}\text{-intrinsic.}$$

Then M' satisfies axiom system D' .

For a proof, see theorem 6.2.2 in Narens (2002a).

Let $\mathcal{F}_{\mathfrak{X}}$ and M' be as in Theorem 9.99. In many situations, each structure in $\mathcal{F}_{\mathfrak{X}}$ can be looked at as a change of context. Then $\mathcal{F}_{\mathfrak{X}}$ -intrinsicness captures not only invariance under symmetries but also invariance under context. A description of this form of invariance under context can be looked at as a way that the contextual invariances produced by the structures in the field are related.

Suppose \mathfrak{X} has the identity as its only symmetry. Then, because axiom system TM holds, every entity in V is M -meaningful. Consider M' to be a meaningfulness predicate that can be used in place of M . Then, in many cases, not every entity in V will be M' -meaningful. The following simple case illustrates this. It is designed to show in a transparent manner how M' can provide a productive theory of meaningfulness for cases where only the identity symmetry is present. Examples with deeper scientific meaning can be found in chapter 6 of Narens (2002a).

- Let $\mathfrak{E} = \langle A, \prec, \oplus \rangle$ be a continuous extensive structure

$$\mathfrak{X} = \mathfrak{X}_a = \langle A, \prec, \oplus, a \rangle, \quad \text{where } a \text{ is a fixed element of } A.$$

- For each b in A , let $\mathfrak{X}_b = \langle A, \prec, \oplus, b \rangle$, and let $\mathcal{F}_{\mathfrak{X}} = \{\mathfrak{X}_b | b \in A\}$.
- Let $\Phi(x, x_1, \dots, x_m)$ be a formula of $\mathbf{L}_{\mathfrak{X}}$, R be an entity in V , p_1, \dots, p_m be pure sets, and for each b in A , the following be a true statement about the structure \mathfrak{X}_b in \mathcal{F} :

$$[\Phi(R, p_1, \dots, p_m) \text{ and for all } x, \Phi(x, p_1, \dots, p_m)] \iff x = R. \quad (9.161)$$

Note that if the symbol a occurs in $\Phi(x, x_1, \dots, x_m)$, then in interpreting Equation (9.161) in \mathfrak{X}_b , the symbol a is interpreted as the element b in A .

Let d be an arbitrary element of A . Because by hypothesis $\mathfrak{E} = \langle A, \prec, \oplus \rangle$ is a continuous extensive structure, it follows that a symmetry α of \mathfrak{E} can be found so that $\alpha(a) = d$. Then:

- (i) α is also an isomorphism of \mathfrak{X}_a onto \mathfrak{X}_d ;
and because for $i = 1, \dots, m$, p_i is a pure set,
- (ii) $\alpha(p_i) = p_i$ for $i = 1, \dots, m$.

From (i), (ii), and Equation (9.161) we derive that R is $\mathcal{F}_{\mathfrak{X}}$ -intrinsic.

Because axiom system TM holds and for each symmetry β of \mathfrak{X}_a , $\beta(a) = a$, it follows from \mathfrak{E} being an extensive structure that β is the identity symmetry, and thus that $M(x)$ holds for each x in V . For each x in V , let $M'(x)$ hold if and only if x is $\mathcal{F}_{\mathfrak{X}}$ -intrinsic. Then by Theorem 9.99, M' is a meaningfulness predicate satisfying axiom system D'. The symmetries of the extensive structure $\mathfrak{E} = \langle A, \prec, \oplus \rangle$ have the following characteristic: for each c and d in A , there is exactly one symmetry γ of $\langle A, \prec, \oplus \rangle$ such that $\gamma(c) = d$. Then by construction, for each entity x in V , $M'(x)$ if and only if x is invariant under the symmetries of \mathfrak{E} .

References

- Aczél, J. (1966). *Lectures on functional equations and their applications*. New York, NY: Academic Press.
- Aczél, J., Roberts, F. S., & Rosenbaum, Z. (1986). On scientific laws without dimensional constants. *Journal of Mathematical Analysis*, 119, 389–416.
- Campbell, N. R. (1920). *Physics: The elements*. Cambridge: Cambridge University Press. Reprinted as *Foundations of science: The philosophy of theory and experiment*. New York, NY: Dover, 1957.
- Campbell, N. R. (1928). *An account of the principles of measurement and calculation*. London: Longmans.
- Cantor, G. (1895). Beiträge zur Begründung der transfiniten Mengenlehre. *Math. Ann.*, 46, 481–512.
- Doble, C. W. (2002). On invariance properties of empirical laws. *Technical Report, Institute for Mathematical Behavioral Sciences, University of California, Irvine*, 02–04.
- Falmagne, J.-Cl. (1971). Bounded versions of Hölder's theorem with application to extensive measurement. *Journal of Mathematical Psychology*, 8(4), 495–507.
- Falmagne, J.-Cl. (1975). A set of independent axioms for positive Hölder systems. *Philosophy of Science*, 42(2), 137–151.
- Falmagne, J.-Cl. (1980). A probabilistic theory of extensive measurement. *Philosophy of Science*, 47, 277–296.
- Falmagne, J.-Cl. (2004). Meaningfulness and order invariance: Two fundamental principles for scientific laws. *Foundations of Physics*, 9, 1341–1384.
- Falmagne, J.-Cl. (2014). On a class of meaningful permutable laws. In C. E. Crangle, A. García de la Sierra, & H.E. Longino (eds.), *Foundations and methods from mathematics to neuroscience: Essays inspired by Patrick Suppes*. Stanford, CA: CSLI Publications.
- Falmagne, J.-Cl. (2015). Deriving meaningful scientific laws from abstract, “gedanken” type, axioms: Three examples. *Aequationes Mathematicae*, 89(2), 393–435.
- Falmagne, J.-Cl. & Doble, C. W. (2015). *On meaningful scientific laws*. Berlin: Springer-Verlag.

- Falmagne, J.-Cl. & Doble, C. W. (2016). Meaningfulness as a “Principle of Theory Construction”. *Journal of Mathematical Psychology*, 75, 59–67.
- Falmagne, J.-Cl. & Doignon, J.-P. (2010). Axiomatic derivation of the Doppler factor and related relativistic laws. *Aequationes Mathematicae*, 80(1), 85–99.
- Falmagne, J.-Cl. & Narens, L. (1983). Scales and meaningfulness of quantitative laws. *Synthese*, 55, 287–325.
- Feynman, R. P., Leighton, R. B., & Sands, M. (1963). *The Feynman lectures on physics*. Reading, MA: Addison-Wesley.
- Helmholtz, H. V. (1887). *Zahlen und messen erkenntnis-theoretisch betrachtet*. Philosophische Aufsätze Eduard Zeller gewidmet, Leipzig, 1887. (Reprinted in Gesammelte Abhandl., 3, 1895, pp. 356–391.) English translation by C. L. Bryan (1930). *Counting and measuring*. Princeton, NJ: Van Nostrand.
- Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen des Königlichen Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physikalische Classe*, 53, 1–64. (Part I translated into English by J. Michell and C. Ernst (1996). “The axioms of quantity and the theory of measurement,” *Journal of Mathematical Psychology*, 40, 235–252.)
- Hosszú, M. (1962a). Note on commutable mappings. *Publ. Math. Debrecen*, 9, 105–106.
- Hosszú, M. (1962b). Néhány lineáris függvényegyenletről. *Mat. Lapok*, 13, 202.
- Hosszú, M. (1962c). Algebrai rendszereken értelmezett függvényegyenletek, i. algebrai módszerek a függvényegyenletek elméletében. *Magyar Tud. Acad. Mat. Fiz. Oszt. Kozl.*, 12, 303–315.
- Kim, S. (1990). On the possible scientific laws. *Mathematical Social Sciences*, 20, 19–36.
- Klein, F. (1872). Vergleichende Betrachtungen über neuere geometrische Forschungen. Verlag von Andreas Deichert, Erlangen, 1872. (Available online at the University of Michigan Historical Mathematics Collection.)
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Volume 1: Additive and polynomial representations*. New York, NY: Academic Press.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1989). *Foundations of measurement, Volume 2: Geometrical, threshold, and probabilistic representations*. New York, NY: Academic Press.
- Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, 66(2), 81–95.
- Luce, R. D. (1962). Comments on Rozeboom’s criticisms of ‘On the Possible Psychophysical Laws’. *Psychological Review*, 69(6), 548–551.
- Luce, R. D. (1964). On a generalization of a theorem of dimensional analysis. *Journal of Mathematical Psychology*, 1, 278–284.
- Luce, R. D. (1967). Sufficient conditions for the existence of a finitely additive probability measure. *Annals of Mathematical Statistics*, 38, 780–786.
- Luce, R. (1978). Dimensionally invariant numerical laws correspond to meaningful qualitative relations. *Philosophy of Science*, 45, 1–16.
- Luce, R. D. (1990). “On the possible psychophysical laws” revisited: Remarks on cross-modal matching. *Psychological Review*, 97(1), 66–77.

- Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement, Volume 3: Representation, axiomatization and invariance*. New York, NY: Academic Press.
- Luce, R. D. & Marley, A. A. J. (1969). Extensive measurement when concatenation is restricted and maximal elements may exist. In S. Morgenbesser, P. Suppes, & M. G. White (eds.), *Philosophy, science and method: Essays in the honor of Ernest Nagel* (pp. 235–249). New York, NY: St. Martin's Press.
- Luce, R. D. & Narens, L. (1985). Classification of concatenation structures by scale type. *Journal of Mathematical Psychology*, 29, 1–72.
- Maksa, G. (2004). CM solutions of some functional equations of associative type. *Annales Univ. Sci. Budapest., Sect. Comp.*, 24, 125–132.
- Maksa, G. (2005). Quasisums and generalized associativity. *Aequationes Mathematicae*, 69, 6–27.
- Narens, L. (1981). A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*, 13, 1–70.
- Narens, L. (1985). *Abstract measurement theory*. Cambridge, MA: MIT Press.
- Narens, L. (1988). Meaningfulness and the Erlanger program of Felix Klein. *Mathématiques Informatique et Sciences Humaines*, 101, 61–72.
- Narens, L. (2002a). *Theories of meaningfulness*. Englewood Cliffs, NJ: Lawrence Erlbaum Associates.
- Narens, L. (2002b). A meaningful justification for the representational theory of measurement. *Journal of Mathematical Psychology*, 46, 746–768.
- Narens, L. (2007). *Introduction to the theories of measurement and meaningfulness and the use of symmetry in science*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Narens, L. (2015). On replacing “quantum thinking” with counterfactual reasoning. E. Dzhafarov, R. Zhang, S. Jordan, & V. Cervantes (eds.), *Contextuality from quantum physics to psychology*. Singapore: World Scientific.
- Narens, L. & Luce, R. D. (1976). The algebra of measurement. *Journal of Pure and Applied Algebra*, 8, 197–233.
- Ng, C. (2016). Functional equations. In W. Batchelder, H. Colonius, E. Dzhafarov, & J. Myung (eds.), *New handbook of mathematical psychology, Vol. 1. Foundations and methodology* (pp. 151–193). Cambridge: Cambridge University Press.
- Osborne, D. (1970). Further extensions of a theorem of dimensional analysis. *Journal of Mathematical Psychology*, 7, 236–242.
- Pfanzagl, J. (1968). *Theory of measurement*. New York, NY: John Wiley & Sons.
- Plateau, M. J. (1872). Sur la mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Académie Royale des Sciences, des Lettres et des Beaux Arts de Belgique*, 33, 376–388, 1872.
- Roberts, F. S. (1979). *Measurement theory, with applications to decisionmaking, utility, and the social Sciences*. Encyclopedia of Mathematics and its Applications, Volume 7. Reading, MA: Addison-Wesley.
- Roberts, F. (1985). Applications of the theory of meaningfulness to psychology. *Journal of Mathematical Psychology*, 29, 311–332.
- Roberts, F. S. & Rosenbaum, Z. (1986). Scale type, meaningfulness, and the possible psychophysical laws. *Mathematical Social Sciences*, 12, 77–95.
- Rozeboom, W. W. (1962a). The untenability of Luce's principle. *Psychological Review*, 69, 532–547.

-
- Rozeboom, W. (1962b). Comment. *Psychological Review*, 69, 552.
- Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, 23, 113–128.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, New Series, 103 (2684), 677–680.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (ed.), *Handbook of experimental psychology* (pp. 1–49). New York, NY: Wiley.
- Suppes, P. (2002). *Representation and invariance of scientific structures*. Stanford, CA: CSLI Publications.
- Suppes, P. & Zinnes, J. (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (eds.), *Handbook of mathematical psychology*, Vol. I. London: John Wiley & Sons.
- Ungar, A. (1991). Thomas precession and its associated grouplike structure. *American Journal of Physics*, 59(9), 824–834.
- Veblen, O. & Young, J. W. (1946). *Projective geometry*, Vol. II. Boston, MA: Ginn and Company.
- Vincze, E. (1962). Eine allgemeinere Methode in der Theorie der Funktionalgleichungen I. *Publ. Math. Debrecen*, 9, 149–163, 1962.

Index

- ($\delta_1, \dots, \delta_n$)-meaningful, 409, 411, 415–417, 441
(numerical) code, 413, 429, 433, 435, 440–441
(positive) relative translation, 429
(quasi-) similarity invariant, 433
1-1-meaningful, 414–415
1-point right solvable, 401
1-point transitive, 438–441
2-code, 401, 403, 404, 410, 411, 414–416, 421
- abstract axioms, 396, 402, 416
 a variant of, 428
 examples, 397
abstract LFD-pair, 406, 418
abstract representation, 399, 402, 403, 405
Aczél, J., 401
admissible transformation, 394
affine equation, 140
AHP, 274, 296, 309, 316
alpha function, 240, 242, 243, 247, 256, 259
analytic, 142, 232, 346, 366
 geometry, 391
anchor, 408
area addition postulate, 419–420
area of the square postulate, 419–420
Arrow’s impossibility theorem, 274
associative, 187, 329, 382, 397, 399, 402, 419
associativity, 299, 397, 402
 Pythagorean theorem, 399, 419–420
atom, 445
attractive force, 413
autodistributive, 405–406
autodistributivity, 397, 405–406
 equations, 398, 405
axiomatic approach, 284, 376
axioms, 105, 196, 201, 202, 218, 274, 285, 290,
 292, 330, 332, 338
 Atomic Legacy, 448
 Meaningful Comprehension, 448
 Meaningful Inheritability, 448
 Meaningful Pure Sets, 448, 449
- backward-graded, 151–152, 167–168, 171, 172,
 174, 178
- Basic Local Independence Model, 132, 162,
 165–166, 181
- Bayesian
 diffusion model, 86, 87, 89, 92
 inference, 84, 87, 101, 325, 355, 360, 363
 models, 100, 101, 355, 359, 360, 362
- Beer’s law, 401, 403, 411, 429, 434–436,
 438–440
- beginning set, 150–151, 160–162
- Bessel process, 48
- binary choice, 2
- birth–death chain, 54
- bisymmetric, 405, 427
 code, 405
- bisymmetry, 397, 405
 equation, 405
- bisymmetry equation, 398, 405
- blood-oxygen-level-dependent (BOLD) signal,
 258
- Boltzmann constant, 404, 441
- Borda assignment rule, 300
- Borda count, 282, 284, 286, 291, 296, 300, 306,
 309, 310, 313, 314
- bounded domain, 428, 429
- Brownian motion, 36, 80, 84, 85
- C-meaningfulness, 437–438
- c-relative translation, 429, 432, 435
 T-class, 432
- canonical form, 21
- cardinality, 150, 166, 196–198
- careless error probability, 153, 168, 173
- Cauchy equation, 404
- Cauchy problem, 169–170
- Cauchy sequence, 189
- Cauchy–Schwarz inequality, 201–202, 333–334
- causality, 112, 116, 126
- centimeter, 396
- characteristic vector, 155–158
- choice probabilities, 19, 62, 65, 138
- choice response time, 2, 27, 69
- class C° , 142
- closure spaces, 151–152

-
- code, 401–406, 408–411
 - numerical, 401
 - solvable, 403
 - commutative, 187, 198, 200, 217, 329, 429, 431
 - family, 429, 431, 434
 - comonotonic, 401, 402, 434–436
 - compartment modeling, 224, 239
 - complex field, 187, 189
 - computational cognitive neuroscience, 224
 - congruence postulate, 419, 420
 - conjugate, 188, 191, 331, 333, 339–341, 430
 - conjugation, 430, 432
 - conjunction fallacy, 204, 205, 217
 - connectionism, 224, 225, 231
 - connectionist models, 224–226, 355, 358
 - continuous extensive structure, 382, 383, 392, 446, 452
 - continuum, 72–74, 391–392
 - convex, 278, 295, 328, 346, 349, 353
 - lower, 429, 432, 433
 - quasi-, 431, 432, 435
 - T-class, 432
 - T-family, 429
 - transformation class, 432
 - upper, 432, 433
 - Correct Response Model, 130, 153, 158, 162
 - counterexample, 147, 403–404, 441
 - critical of rank k , 144
 - cylinder, 403
 - volume, 412
 - decimeter, 396
 - decision boundary, 36, 353
 - decision criterion, 36, 74
 - decision-bound models, 355, 357
 - decision-making, 2, 4, 5, 7, 8, 11, 31, 33, 36, 49, 58, 72, 74, 76, 80, 82, 83, 87, 89, 186, 204, 285, 292, 324
 - Dedekind complete, 392
 - dense, 38, 83, 242, 333, 351
 - density matrix, 213
 - denumerable density, 392
 - diffeomorphism, 143
 - differentiable, 51, 141, 346, 349
 - differential, 141
 - diffusion
 - coefficient, 39, 42, 50, 52, 59
 - model, 2–4, 71, 80, 82–84, 89, 92, 95, 97, 98, 100, 101, 109
 - process, 38, 39, 49, 50, 58, 82–92, 100, 257, 357
 - rate, 82, 84, 86–89, 91, 94, 99, 100
 - dimensional constant, 395, 412, 442
 - dimensionally invariant, 415
 - dimensionally order-invariant, 415
 - discriminative, 149, 150
 - disjunction effect, 207, 217
 - domain, 129–131, 133, 149, 429
 - of a family, 429
 - pertinent, 430
 - dopamine, 226, 240, 245–252
 - drift criterion, 85–87, 90, 102
 - drift rate, 39, 41, 49, 59, 77, 78, 82, 84–92, 98–102
 - dual, 151–152, 167, 188–190, 330–332, 334, 336, 339, 341, 348, 351–352, 367
 - duality mapping, 328, 329, 331, 333–335, 339, 341
 - edges, 30, 140, 276, 306
 - endomorphism, 386–387
 - entity, 445, 447, 449, 452–453
 - purely mathematical, 446
 - scientific, 446, 451
 - equally informative, 171–172
 - equivalence relation, 151, 383, 430
 - Erlanger program, 375, 389, 393, 443–449
 - errors
 - fast, 98, 101
 - slow, 98, 101
 - evidence sampling, 71, 100
 - exemplar models, 326, 328, 354, 355, 357, 358
 - exponential decay, 404, 428
 - extension, 151
 - face, 140, 276–278, 306, 344, 356
 - facets, 140
 - fair, 18, 55, 153, 154, 157, 163–172, 274, 275, 284
 - family of (numerical) codes, 413
 - feature space, 325–328, 344, 346, 348–350, 352–354, 359, 362–367
 - filtration, 111, 113, 116, 117
 - firing rate model, 241–244
 - firm, 151
 - first exit time, 19, 113
 - first passage time, 19, 23, 27, 56, 61, 62
 - first-order relation, 446
 - fixed stopping rule, 5
 - fMRI, 224, 226, 227, 254, 257–261
 - model-based, 258
 - forward-graded, 151–152, 167, 174
 - forward-graded in item, 151
 - full, 140
 - full rank, 142, 173–175
 - functional equation, 395
 - fundamental matrix, 22
 - G*-permutable, 402
 - Gaussian vector, 36, 46–48
 - geometric means approach, 296
 - geometric profile representation, 281, 285

- gravitational constant, 379, 381, 412
 gravitational force, 412
 group, 210, 211, 316, 317, 329
 invariant of, 390
 of symmetries, 389–391, 449, 451
 of transformations, 389–391, 447
- heavy notation, 396
 Hebbian learning, 245, 247–249, 252, 253
 higher-order, 390, 391, 443–446
 relations, 390, 444–446
 Hilbert space, 187–192, 194, 195, 197, 199–201,
 203, 204, 218, 324, 327, 331–333, 335,
 336, 338–342, 344, 346, 348, 350, 353,
 363, 367
 hitting time, 19
 Hodgkin–Huxley model, 231
 homogeneity, 330, 401, 402, 404, 416
 homogeneous, 400, 401, 420, 425, 427, 441
 code, 401
 time-, 10, 34, 62
 homomorphism, 384, 393
 hypotenuse, 396, 397, 399, 402, 404, 419,
 425–426
- identifiable, 133–139, 146–148
 at a parameter point, 133
 at a predicted point, 133
 imperfect discrimination, 292, 293, 324
 incidence matrix, 155–156, 158–159
 initial code, 409, 418, 420, 425–429
 initial unit, 396, 412
 inner product, 187–189, 192, 193, 195, 197, 201,
 328, 329, 331–333, 335, 336, 338, 340,
 343, 348, 349, 363
 intensity, 210, 214, 215, 313–315, 376–377, 403,
 440
 interference, 205, 213, 214, 216, 217
 intertrial variability, 97, 98, 101, 102
 interval scale, 90, 376, 377, 396
 intransitive rankings, 276, 277
 intrinsic
 \mathcal{F} -intrinsic, 452–453
 invariant, 90, 99, 100, 121, 230, 231, 238, 390
 isotone, 415–416
 item, 149
 Izhikevich model, 235–238
- Klein, 375, 388–389, 443, 447
 knowledge space, 150, 152
 knowledge state, 129, 149, 152–153, 162, 164,
 172, 173
 knowledge structure, 135, 148–156, 158–159,
 162, 164, 166, 168–172, 174–175, 178,
 181
 Kruskal–Wallis, 290–291, 300, 305
- Law of universal gravitation, 395
 leaky integrate-and-fire model, 232–234
 learning space, 149–152, 172
 left homogeneity, 404
 left meaningful, 436
 left order-invariant, 433, 434, 441–442
 legitimate left operator, 438–440
 legitimate right operator, 437–439
 length, 396
 LF-contraction, 434, 438–439
 order-invariance, 400
 standard form, 438
 LFD pair
 abstract, 406, 418
 LFD system, 417–419, 428
 linear description, 140, 156
 linear equation, 140
 local field potentials, 259
 locally identifiable, 137–139, 142, 146–147
 at a point, 137
 locally injective, 140, 142
 log-likelihood ratio, 75–77, 80, 83, 85, 86, 88, 89
 log-odds, 75–77, 79, 81, 83–85, 87–90, 92, 94
 long-term depression (LTD), 245
 long-term potentiation (LTP), 245
 Lorentz–FitzGerald
 abstract representation, 398–399
 contraction, 399–401, 403, 407, 408, 410,
 411, 439
 Lorentz–FitzGerald–Doppler
 system, 417
 equation, 398, 406
 abstract, 398, 406
 lower convex, 429, 431, 433
 lower convex T-class, 432
 lower convex T-family, 429
 Luce’s choice axiom, 292, 293
 lucky guess probability, 153
- Markov
 chain, 3, 8–18, 20, 21, 23, 24, 28, 30, 33,
 59–63
 model, 4, 212, 214, 215, 217
 bounded memory, 17
 process, 4, 33, 35, 42, 49, 50, 52, 53, 62, 77,
 208
 property, 4, 9, 15, 18, 24, 82
 martingale, 54–57
 mass, 379, 382, 383, 396, 412
 meaningful, 395, 439, 440, 443
 left, 436, 438
 right, 436, 438
 meaningful LFD system, 417, 418
 meaningfulness, 410, 416
 definition, 409
 meaningless, 443, 450

-
- measurement operator, 203, 212, 215
 measurement theory, 376, 394, 434
 mental architecture, 106
 meter, 396
 model, 133
 monomial laws, 429, 436, 438, 440
 monotonicity
 extensive structure, 382
 multiplicative representation, 434
- n*-code, 401
n-dimensional (numerical) code, 401
 Nakamura number, 312
 Narens, 379, 387, 393, 443, 449, 450
 neural networks, 186, 358
 neuroscience ideal, the, 230
 Newton's law of gravity, 395
 non-negative, 10, 18, 33, 43, 401, 429
 non-parametric statistics, 272, 290, 300
 norm, 189, 324, 327, 328, 330–333, 335, 338,
 341, 344–347, 350–352, 366
 nullcline, 238
- one-to-one meaningful, 414–415
 one-to-one meaningfulness, 412
 optional stopping rule, 5
 order schedule, 59–63, 65, 67, 68
 order-invariance, 400, 428, 433, 437, 442
 order-meaningful, 413–416
 order-meaningfulness, 412–415
 ordinal space, 150, 162
 Ornstein–Uhlenbeck process, 49, 51–54, 59, 241
 outcome point, 133–135, 139, 153, 154
 outcome space, 133–138, 147, 154, 164, 177,
 179–181, 284, 285
- paired comparison difficulties, 296, 299, 314
 parameter, 135
 parameter domain, 130, 133–139, 145–147,
 157–163, 167–172
 parameter point, 133–138, 141, 153–154,
 158–161, 167–169, 178–180
 parameter space partitioning, 263
 parameterized probabilistic knowledge structure,
 153
 partial order, 150, 162
 path dependency, 296, 299, 309, 314
 permutability, 397, 402–403, 416, 425–427
 permutable, 402–404, 417, 420–421, 423, 427
 with respect to, 402, 421–423, 427
 Pexider equation, 422
 Pfanzagl, J., 385
 phase portrait, 233–237
 polytope, 140, 156, 158
 of the knowledge structure, 156
 0/1, 140
- positional voting rules, 284, 293
 positive operator valued measurement, 203
 positivity, 209, 330, 331, 333, 382
 predicted point, 131–135, 137, 153–154,
 161–162, 167
 prediction function, 133–134, 136–137,
 139–142, 146–148
 prediction range, 130, 133–134, 136, 137,
 155–159, 161–162, 179
 pressure, 396, 404, 441
 principle of correspondent change, 229–230
 principle of theory construction, 378, 394
 probabilistic, 133
 probabilistic knowledge structure, 153, 178
 processing order, 60, 106, 121, 122
 projection operator, 217
 prototype models, 354, 356, 361
 psychophysical laws (possible), 377
 purely mathematical entity, 446
 Pythagorean theorem, 395, 397, 400, 401, 404,
 412, 419, 420, 425
- quadratic integrate-and-fire model, 234–235
 qualitatively \mathcal{S} -meaningful, 385–387
 quantitatively testable, 136–137, 139, 146, 148,
 156, 165–166, 179, 180
 quantum cognition, 186, 189, 196
 quantum probability theory, 200–202
 quantum thinking, 274, 309, 314
 quasi-*c*-relative translation, 433, 435, 437
 quasi-arithmetic mean, 405
 quasi-convex, 431, 432, 435
 lower, 431
 upper, 431
 quasi-convex invariant, 435
 lower, 435
 quasi-lower convex, 437, 433–437
 quasi-order, 150–151, 160–162
 quasi-ordinal space, 150, 161–162
 quasi-permutability, 397, 402–403, 419, 420,
 425–427
 quasi-permutable, 402, 403, 420, 427
 quasi-similarity, 431
 quasi-upper convex, 431, 433, 435–437, 439
- radiation, 403, 439
 random walk, 2–5, 8, 12, 18, 24–32, 35–37, 41,
 45, 57, 62, 75, 77, 79–84, 88, 100, 213,
 356
 rank, 142
 rank-deficient, 142, 168, 169
 ranking wheel configuration, 307, 308, 311–313
 rate constant, 404
 ratio scale, 139, 274, 316, 376–377, 382–383,
 395–397, 401, 407, 414–416, 430
 ratio scale matrix, 316

- reducible on both sides, 401–402, 406
 regular, 143–144, 146–148, 236–237, 341
 reinforcement learning, 245, 246, 248–251, 258,
 262, 324, 325, 365
 relative neighborhood, 137–138
 representer theorem, 329, 347, 348, 351, 354,
 367
 reproducing kernel, 324, 327–329, 338–340,
 342, 344, 346–353, 361, 362, 365, 367
 response probability, 71, 92, 94–96, 98, 158
 response time distribution, 27, 108, 110
 reward prediction error, 250–252
 Riesz representation, 332–334, 338
 right meaningful, 436–437
 right order-invariant, 433, 435–437, 442–443
 right triangle, 279, 281, 396, 399, 402, 404, 420,
 425–426
 Roberts, 378, 385, 395

 scaling degeneracy, 89–92
 scientific entity, 446, 451
 scientific law (possible), 394, 400, 404
 scientific topic, 443–451
 principles, 444
 Scott, 384
 selective influence, 108, 114, 120–126
 self-transforming, 401, 402, 410, 417
 self-transforming collections, 401
 definition, 410
 semi-inner product, 328, 329, 333–336,
 338–341, 349, 350, 352–354, 363, 364,
 366, 367
 serial processing, 115
 set of atoms, 445
 set-in-stone ideal, 231
 signal detection model, 72–77, 100, 214
 signal-to-noise ratio, 86, 87, 102
 similarity, 13, 33, 107, 324, 326–328, 346, 348,
 349, 351, 354–357, 359, 361–367, 426,
 428–430, 433
 similarity family, 429, 430, 433–435, 437–438
 similarity invariant, 433
 similarity transformation, 428
 simplex, 154–156, 159–162, 279
 simplicity heuristic, 230–231
 single-unit recordings, 236, 254, 257
 solvability, 382, 383, 398, 417
 extensive structure, 382
 solvable, 401, 403, 417
 code, 401, 403, 417
 sparsity, 324, 327, 328, 343, 346, 352, 361, 362,
 367
 speed–accuracy trade-off, 79, 87, 98
 spike-timing-dependent plasticity, 253
 spiking neuron models, 231, 238, 239, 241, 242,
 254
- Hodgkin–Huxley, 231
 Izhikevich, 234–239, 241, 254–255
 leaky integrate-and-fire, 232–234, 241, 243
 quadratic integrate-and-fire, 234–236, 238,
 257
 ST-collection, 410, 412, 417, 419, 427, 428
 starting-point variability, 98
 state
 absorbing, 6, 14, 21, 22, 27, 58, 63
 transient, 6, 22
 stochastic differential equation, 39, 49, 53, 84,
 241
 stochastic matrix, 10, 12, 63, 64, 210
 stochastic process, 2–5, 17, 18, 33, 49, 54, 55,
 62, 65, 76, 82, 84, 85, 105, 108,
 110–112, 116, 125
 stopping time, 5, 11, 18–20, 42–45, 56, 57, 112
 straight, 153–162, 165
 striatum, 226, 228, 245, 246, 248, 263
 strong transitivity, 298, 302, 317
 subspace, 140, 142, 144–145, 159, 189–194,
 197, 204, 217, 286, 288–290, 294, 298,
 305, 307, 330–334, 341, 350
 supermajority voting, 311
 superposition, 209, 214, 217
 supervised learning, 245
 Suppes, 138, 196, 378, 383–385, 393, 395, 400
 supporting hyperplane, 140
 symmetric, 37, 48, 74, 90–92, 150, 330, 334,
 339, 346, 401, 420
 code, 401, 403, 406
 symmetric difference, 150
 symmetry, 388
 group of, 389, 450
 synaptic plasticity, 226, 244–248, 250, 261
 synthetic geometry, 392

 T-class, 432, 436–440
 T-family, 429–430, 433, 436–440
 commutative, 431
 transitive, 430, 431, 438–439, 441
 T-meaningfulness, 431–433, 437–440
 testable, 133–137, 139, 146, 148, 156–157,
 165–166, 179–180, 380
 tight, 151, 160
 time, 396
 time schedule, 60, 64, 68
 time-homogeneous, 10, 34, 62
 time-invariant, 259
 trajectory, 81, 83–85, 118, 169, 209, 238
 transcranial magnetic stimulation, 224, 261–262
 transformation
 class, 428, 431
 classes, 431
 convex, 428, 431
 family, 429, 431

-
- meaningful, 428, 436
 - similarity, 428
 - transformation group, 389, 391, 393, 447
 - of a structure, 389, 393
 - transition matrix, 10, 12, 13, 15, 16, 21, 34, 50, 58, 61, 63, 209, 210, 214–215
 - transition probability, 7, 10, 15, 50, 65
 - transitive, 274, 276, 277, 429, 431
 - family, 429, 431, 434, 438–440
 - translatable, 404, 405, 417, 428
 - translation equation, 397, 404–405
 - triangulated, 162

 - u*-conjugate, 430, 432
 - unfalsifiability, 97, 98, 101
 - union closed, 150
 - unit
 - initial, 396, 408, 412
 - unitary operator, 191–195
 - unsupervised learning, 244
 - upper convex T-class, 432–433, 436, 438–439
 - upper convex T-family, 429

 - van der Waals' equation, 404, 429, 441, 443

 - Veblen, 447
 - vector space, 186–190, 199–201, 208, 326–333, 335–338, 340, 341, 344, 347, 348, 363
 - velocity, 51, 406
 - vertex, 31, 140, 155–156, 159, 162, 276, 277, 279–282, 306
 - Vincze's equation, 422
 - voting theory, 272–274, 276, 285, 290–294

 - weak order, 150–151, 160–161
 - well-graded, 150, 152
 - white noise, 37, 41, 241
 - Wiener process, 4, 19, 35–46, 48–51, 53–55, 57–59, 83
 - multidimensional, 46
 - standard, 36, 38, 39, 41, 43, 49
 - with drift, 39–42, 44, 46, 57

 - x_0 -solvable, 401
 - for codes, 401

 - Young, 447

 - Zinnes, 138, 385

The field of mathematical psychology began in the 1950s and includes both psychological theorizing, in which mathematics plays a key role, and applied mathematics motivated by substantive problems in psychology. Central to its success was the publication of the first *Handbook of Mathematical Psychology* in the 1960s. The psychological sciences have since expanded to include new areas of research, and significant advances have been made in both traditional psychological domains and in the applications of the computational sciences to psychology. Upholding the rigor of the original handbook, the *New Handbook of Mathematical Psychology* reflects the current state of the field by exploring the mathematical and computational foundations of new developments over the last half-century. The second volume focuses on areas of mathematics that are used in constructing models of cognitive phenomena and decision-making, and on the role of measurement in psychology.

WILLIAM H. BATCHELDER
is Professor of Cognitive Sciences at the University of California, Irvine.

HANS COLONIUS is Professor of Psychology at Oldenburg University, Germany.

EHTIBAR N. DZHAFAROV is Professor of Psychological Sciences at Purdue University.

“This handbook contains review articles by leaders of the field. It will be a valuable reference for all who are interested in mathematical psychology.”

BRIAN SKYRMS, *University of California, Irvine*

“This volume is the second in a series devoted to presenting a very readable, yet broad and deep, up-to-date perspective on mathematical psychology. It will be of interest to graduate students and experienced researchers who are concerned with modeling and measurement in the psychological and cognitive sciences.”

THOMAS S. WALLSTEN, *University of Maryland*

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-1-107-02907-1



9 781107 029071 >