Handbook of Philosophical Logic 18

Dov M. Gabbay
Franz Guenthner *Editors*

Handbook of Philosophical Logic

Volume 18

Springer

Handbook of Philosophical Logic

HANDBOOK OF PHILOSOPHICAL LOGIC

Volume 18

edited by Dov M. Gabbay and Franz Guenthner

Volume 1 – ISBN 0-7923-7018-X
Volume 2 – ISBN 0-7923-7126-7
Volume 3 – ISBN 0-7923-7160-7
Volume 4 – ISBN 1-4020-0139-8
Volume 5 – ISBN 1-4020-0235-1
Volume 6 – ISBN 1-4020-0583-0
Volume 7 – ISBN 1-4020-0599-7
Volume 8 – ISBN 1-4020-0665-9
Volume 9 – ISBN 1-4020-0699-3
Volume 10 – ISBN 1-4020-1644-1
Volume 11 – ISBN 1-4020-1966-1
Volume 12 – ISBN 1-4020-3091-6
Volume 13 – ISBN 978-1-4020-3520-3
Volume 14 – ISBN 978-1-4020-6323-7
Volume 15 – ISBN 978-94-007-0484-8
Volume 16 – ISBN 978-94-007-0478-7
Volume 17 – ISBN 978-94-007-6599-3

More information about this series at http://www.springer.com/series/6024

Dov M. Gabbay • Franz Guenthner
Editors

Handbook of Philosophical Logic

Volume 18

Springer

Editors
Dov M. Gabbay
Department of Computer Science
King's College London
London, United Kingdom

Franz Guenthner
Centrüm für Informations- und Sprachvera
University of Munich
Munich, Germany

Handbook of Philosophical Logic
ISBN 978-3-319-97754-6 ISBN 978-3-319-97755-3 (eBook)
https://doi.org/10.1007/978-3-319-97755-3

Library of Congress Control Number: 2018957613

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

CONTENTS

PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as 'the best starting point for exploring any of the topics in logic'. We are confident that the second edition will prove to be just as good!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and λ-calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and λ-calculus was too far removed.[1] Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

[1] I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract an similar to principles governing the cooperation of two large organisation. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors are readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay
King's College London, and
Bar Ilan University, Israel, and
University of Luxembourg

Logic	**IT**			
	Natural language processing	**Program control specification, verification, concurrency**	**Artificial intelligence**	**Logic programming**
Temporal logic	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
Modal logic. Multi-modal logics	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
Algorithmic proof	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
Non-monotonic reasoning	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
Probabilistic and fuzzy logic	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
Intuitionistic logic	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
Set theory, higher-order logic, λ-calculus, types	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	λ-calculus extension to logic programs

Imperative vs. declarative languages	**Database theory**	**Complexity theory**	**Agent theory**	**Special comments: A look to the future**
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

Classical logic. Classical fragments	Basic background language	Program synthesis	A basic tool	
Labelled deductive systems	Extremely useful in modelling		A unifying framework. Context theory.	Annotated logic programs
Resource and substructural logics	Lambek calculus		Truth maintenance systems	
Fibring and combining logics	Dynamic syntax	Modules. Combining languages	Logics of space and time	Combining features
Fallacy theory				
Logical Dynamics	Widely applied here			
Argumentation theory games		Game semantics gaining ground		
Object level/ metalevel			Extensively used in AI	
Mechanisms: Abduction, default relevance			ditto	
Connection with neural nets				
Time-action-revision models			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model

JOHAN VAN BENTHEM AND FENRONG LIU

DEONTIC LOGIC AND CHANGING PREFERENCES

ABSTRACT: The normative realm involves deontic notions such as obligation or permission, as well as information about relevant actions and states of the world. This mixture is not static, given once and for all. Both information and normative evaluation available to agents are subject to changes with various triggers, such as learning new facts or accepting new laws. This paper explores models for this setting in terms of dynamic logics for information-driven agency. Our paradigm will be dynamic-epistemic logics for knowledge and belief, and their current extensions to the statics and dynamics of agents' preferences. Here the link with deontics is that moral reasoning may be viewed as involving preferences of the acting agent as well as preferences of moral authorities such as lawgivers, one's conscience, or yet others. In our presentation of preference based agency, we discuss a large number of themes: primitive 'betterness' order versus reason-based preferences (employing a model of 'priority graphs'), the entanglement of preference and informational attitudes such as belief, interactive social agents, and scenarios with long-term patterns emerging over time. Specific deontic issues considered include paradoxes of deontic reasoning, acts of changing obligations, and changing norm systems. We conclude with some further directions, such as multi-agency and games, plus pointers to related work, including different paradigms for looking at these same phenomena.

1 AGENCY, INFORMATION, AND PREFERENCE

Agents pursue goals in this world, acting within constraints in terms of their information about what is true, as well as norms about what is right. The former dimension typically involves acts of inference, observation, as well as communication and other forms of social interaction. The latter dimension involves evaluation of situations and actions, 'coloring' the agents' view of the world, and driving their desires, decisions, and actions in it. A purely informational agent may be rational in the sense of clever reasoning, but a *reasonable* agent is one whose actions are in harmony with what she wants. The two dimensions are intimately related. For instance, what we want is influenced by what we believe to be true as well as what we prefer, and normally also, we only seek information to further goals that we desire.

This balance of information and evaluation is not achieved once and for all. Agents must constantly cope with new information, either because they learn more about the current situation, or because the world has changed. But equally well, agents constantly undergo changes in evaluation, sometimes by intrinsic changes of heart, but most often through events with normative impact, such as accepting a command from an authority. These two forms of dynamics, too, are often entangled: for instance, learning more about the facts can change my evaluation of a situation.

A third major aspect of agency is its social interactive character. Even pure information flow is often driven by an epistemic gradient: the fact that different agents know different things leads us to communicate, whether in

D. M. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*,
Handbook of Philosophical Logic 18, https://doi.org/10.1007/978-3-319-97755-3_1

cooperative inquiry or adversarial argumentation, perhaps until a state of equilibrium is reached such as common knowledge or common belief. But also more complex forms of interaction occur, such as merging beliefs in groups of agents, where differences in informational authority may play a crucial role. Again, very similar phenomena play on the normative side. Norms, commitments and duties usually involve other agents, both as their source and as their target, and whole institutions and societies are constructed in terms of social choice, shared norms and rules of behavior.

In this chapter, we take current "dynamic-epistemic logics" as our model for the above phenomena, informational and preferential, and we show how this perspective transfers to normative reasoning and deontic logic. We will highlight two main themes: (a) making the dynamic *actions and events* that drive real deontic scenarios, such as commands or permissions, an explicit part of our analysis, and (b) exploring more finely structured *reasons for deontic preferences*. Important side-themes linked with these are (c) the entanglement of obligation, information, knowledge and belief, and (d) the importance of multi-agent scenarios, such as games, in the deontic realm. Our treatment will be brief, and for a more elaborate sample of this style of thinking about the normative realm, we refer to [Benthem *et al.*, 2014].

In pursuing the specific technical paradigm of this chapter, we are not at all denying the existence of other valid approaches to deontic dynamics or further themes covered, and we will provide a number of references to other relevant strands in the literature.

2 DYNAMIC LOGICS OF KNOWLEDGE AND BELIEF CHANGE

Before analyzing preference or related deontic notions, we first develop the basic methodology of this paper for the purely informational case, where the first dynamic-epistemic logics arose in the study of information update and information exchange between agents.

2.1 Epistemic logic and semantic information

Dynamic logics of agency need an account of underlying of static states that can be modified by suitable triggers: actions or events. Such states usually come from existing systems in philosophical or computational logic whose models can serve as static snapshots of the dynamic process.

In this chapter, we start with a traditional modal base system of epistemic logic, referring to the standard literature for details (cf. [Fagin *et al.*, 1995] and [Blackburn *et al.*, 2001]).

Definition 1 *Let a set of propositional variables Φ be given, as well as a set of agents A. The epistemic language is defined by the syntax rule*

$$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \qquad \textit{where } p \in \Phi,\ a \in A.$$

Remark: Single agents, interacting agents, and groups. For convenience, we will focus on single agents in this chapter – something that still allows us to describe interacting individual agents where needed through iterations of modalities. Epistemically important notions with groups themselves as agents, such as 'common knowledge' or 'distributed knowledge', are deferred to our discussion at the end. Group actors are also important in the deontic realm, involving collective commitments or duties, but we will only touch upon this theme occasionally.

Semantic models for the epistemic language encode agents' 'information ranges' in the form of equivalence classes of binary uncertainty relations for each agent.[1] These support a standard compositional truth definition.

Definition 2 *An* epistemic model *is a tuple* $\mathfrak{M} = (W, \{\sim_a\}_{a\in A}, V)$ *with* W *a set of epistemically possible states (or 'worlds'),* $\sim_a$ *an equivalence relation on* W*, and* V *a valuation function from* Φ *to subsets of* W*.*

Definition 3 *For an epistemic model* $\mathfrak{M} = (W, \{\sim_a \mid a \in A\}, V)$ *and any world* $s \in S$*, we define* $\mathfrak{M}, s \models \varphi$ *(epistemic formula* φ is true in $\mathfrak{M}$ at s*) by induction on the structure of the formula* φ*:*

1. $\mathfrak{M}, s \models \top$ *always.*
2. $\mathfrak{M}, s \models p$ iff $s \in V(p)$.
3. $\mathfrak{M}, s \models \neg\varphi$ iff *not* $\mathfrak{M}, s \models \varphi$.
4. $\mathfrak{M}, s \models \varphi \wedge \psi$ iff $\mathfrak{M}, s \models \varphi$ *and* $\mathfrak{M}, s \models \psi$.
5. $\mathfrak{M}, s \models K_a\varphi$ iff *for all* t *with* $s \sim_a t : \mathfrak{M}, t \models \varphi$.

Using equivalence relations in our models yields the well-known modal system **S5** for each individual knowledge modality, without interaction laws for different agents. Just for concreteness, we state this basic fact here:

Theorem 4 *Basic epistemic logic is axiomatized completely by the axioms and inference rules of the modal system* ***S5*** *for each separate agent.*

Few researchers see our basic modalities and the simple axioms of modal **S5** as expressing genuine properties of 'knowledge' – thus making earlier polemical discussions of epistemic 'omniscience' or 'introspection' expressed

[1] The approach of this paper will also work on models with more general relations such as transitive and reflexive pre-orders, but we start with this easily visualizable epistemic case for expository purposes.

by these axioms obsolete. Our interpretation of the above notions is as describing the *semantic information* that agents have available (cf. [Benthem, 2014]), being a modest but useful building block in analyzing more complex epistemic and deontic notions. We will allow ourselves the use of the word 'know' occasionally, however: old habits die hard. [2]

Now comes our first key theme. Static epistemic logic describes what agents know on the basis of their current semantic information. But information flows, and a richer story must also include dynamics of actions that produce and modify information. We now turn to the simplest case of this dynamics: reliable public announcements or public observations, that shrink the current information range.

2.2 Dynamic logic of public announcement

The pilot for the methodology of this paper is 'public announcement logic' (PAL), a toy system describing a combination of epistemic logic and one dynamic event, namely, *announcement* of new 'hard information' expressed in some proposition φ that is true at the actual world. The corresponding 'update action' $!\varphi$ transforms a current epistemic model $\mathfrak{M}, s$ into its definable submodel $\mathfrak{M}|\varphi, s$ where all worlds that did not satisfy φ have been eliminated. This model update is the basic scenario of obtaining information in the realm of science but also of common sense, by shrinking one's current epistemic range of uncertainty.[3]

To describe this phenomenon, the *language* of PAL has two levels, using both formulas for propositions and action expressions for announcements:

$$\begin{aligned}&\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \mid [A]\varphi\\&A := !\varphi\end{aligned}$$

The new dynamic formula $[\varphi]\psi$ says that "after updating with the true proposition φ, formula ψ holds":

$$\mathfrak{M}, s \models [!\varphi]\psi \;\text{ iff }\; \text{if } \mathfrak{M}, s \models \varphi, \text{ then } \mathfrak{M}|\varphi, s \models \psi.$$

This language can make characteristic assertions about knowledge change such as $[!\varphi]K_a\psi$, which states what agent a will know after having received the hard information that φ. In particular, the knowledge change before and after an update can be captured by so-called *recursion axioms*, a sort

[2]There is a fast-growing literature on more sophisticated logical analyses of genuine knowledge (cf. [Holliday, 2012], [Benthem and Pacuit, 2011], [Shi, 2014]), which also seems relevant to modeling and reasoning in the deontic realm. However, the main points to be made in this chapter are orthogonal to these additional refinements.

[3]The name 'public announcement logic' may be unfortunate, since the logic PAL describes updates with hard information from whatever source, but no consensus has emerged yet on a rebaptism.

of recursion equations for the 'dynamical system' of PAL, relating new knowledge to knowledge that agents had before. Here is the complete logical system for information flow under public announcement (two original sources are [Gerbrandy, 1999], [Plaza, 1989]):

Theorem 5 *PAL is axiomatized completely by the usual laws of the static epistemic base logic plus the following recursion axioms:*

1. $[!\varphi]q \leftrightarrow (\varphi \to q)$ *for atomic facts q*
2. $[!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$
3. $[!\varphi](\psi \wedge \chi) \leftrightarrow ([!\varphi]\psi \wedge [!\varphi]\chi)$
4. $[!\varphi]K_a\psi \leftrightarrow (\varphi \to K_a[!\varphi]\psi)$

These elegant principles analyze reasoning about epistemic effects of receiving hard information, through observation, communication, or other reliable means. In particular, the knowledge law reduces knowledge after new information to 'conditional knowledge' that the agent had before, but in a subtle recursive manner. This prudence of design for PAL is necessary since the dynamic process of information update can typically change truth values of epistemic assertions over time. Perhaps, initially, I did not know that p, but after the event $!p$, I do.

There are several noteworthy features to this approach. We already stressed the recursive nature of reducing new knowledge to pre-existing knowledge, a feature that is typical of dynamical systems. Also, the precise way in which this happens involves breaking down, not the announced propositions (as one might expect), but the 'postconditions' behind the dynamic modalities $[!\varphi]$ compositionally on the basis of their shape. [4]

Next, as things stand here, repeating these steps, the stated features drive a 'reduction process' taking every formula of our dynamic-epistemic language eventually to an equivalent formula inside the static epistemic language. In terms of semantics and expressive power, this means that a current static model 'pre-encodes' all information about what might happen when agents communicate what they know. In terms of the logic, the reduction procedure means that PAL is *axiomatizable* and *decidable*, since it inherits these features from the epistemic base logic.

However, it is important to note that sweeping dynamics-to-statics reduction is not an inevitable feature of dynamic-epistemic analysis. In recent semantics for PAL, available sequences of updates are constrained by *protocols* that restrict available events in the current process of inquiry. In that case, no reduction is possible to the base logic, and the dynamic logic,

[4]One can have compound informational actions, of course, but these would rather be modeled in an extended syntax of programs over atomic announcement actions.

though still employing recursion equations, and remaining axiomatizable and decidable, comes to encode a genuine new kind of 'procedural information' (cf. [Benthem *et al.*, 2009a]). Protocols also make sense for deontic purposes, because of the procedural character of much normative behavior, and we will briefly return to this perspective at the end of this chapter.

In what follows, *PAL* will serve as a pilot example for many other complex cases, for example, changes in beliefs, preferences, and obligations. In each case, the 'triggering events' can be different: for instance, beliefs can change by signals of different force: hard or more 'soft', and obligations can change through actions of commanding by a normative authority. In many cases, the domain of the model does not change, but rather its *ordering pattern*.[5] However, the general recursive methodology of *PAL* will remain in force, though in each case, with new twists.

2.3 From knowledge to belief

Knowledge rests on hard information, but most of the information that we have and act on is soft, giving rise to *beliefs*, that are not always true, and that can be revised when shown inadequate. One can think of learning from error as the more creative ability, beyond mere recording of reliable information in the agent's environment.

Again we need to start with a convenient static base for our investigation. One powerful model for soft information and belief reflects the intuition that we believe those things that hold in the *most plausible* worlds in our epistemic range. I believe that this train will take me home on time, even though I do not know that it will not suddenly fly away from the tracks. But the worlds where it stays on track are more plausible than those where it flies off, and among the latter, those where it arrives on time are more plausible than those where it does not.

The long history for this way of modeling belief includes non-monotonic logic in artificial intelligence ([Shoham, 1988], [Boutilier, 1992], [Lamarre and Shoham, 1994], [Friedman and Halpern, 1997], [Friedman and Halpern, 1999]), [6] the semantics of natural language (cf. [Veltman, 1996]), as well as the philosophical literature on epistemology and games (cf. [Stalnaker, 1996], [Baltag and Smets, 2008]). The common intuition of relative plausibility leads to the following semantics:

[5] One example of this approach, even in the epistemic realm, are 'link cutting' versions of updating after announcement: cf. [Liu, 2004], [Snyder, 2004], [Benthem and Liu, 2007]. Such transformations will be used later on in scenarios where we may want to return to worlds considered earlier in the process.

[6] More generally, non-monotonic logic has been a continuing source of inspiration for logics of belief, preference, and even deontic logic – for instance, in the treatment of conditional belief or conditional obligation. As another type of illustration, the two last-mentioned papers also show analogies with our Section 6 on reasons for preference.

Definition 6 *An* epistemic-doxastic model $\mathfrak{M} = (W, \{\sim_a\}_{a\in A}, \{\leq_a\}_{a\in A}, V)$ *consists of an epistemic model* $(W, \{\sim_a\}_{a\in A}, V)$ *as before, while the* $\leq_a$ *are binary comparative plausibility pre-orders for agents between worlds.*

Intuitively, these comparison orders might well be *ternary* $\leq_{a,s} xy$ saying that, in world s, agent a considers world x at least as plausible as y. [7] For convenience in this chapter, however, our semantics assumes that plausibility orderings are the same for epistemically indistinguishable worlds: that is, agents know their plausibility judgements. Assuming that plausibility is a pre-order, i.e., reflexive and transitive, but not necessarily connected, leaves room for the existence of genuinely incomparable worlds – but much of what we say in this chapter also holds for the special case of *connected* pre-orders where any two worlds are comparable.[8]

As with epistemic models, however, our style of logical analysis will work largely independently from specific design decisions about the ordering, important though these may be in specific applications.

One can interpret many logical languages in these comparative order structures. In what follows, we will mainly work with modal formalisms – for the usual reasons of perspicuous formulation and low complexity (cf. [Blackburn *et al.*, 2007]).

First of all, there is *absolute belief* as truth in all most plausible worlds:

$\mathfrak{M}, s \models B_a\varphi$ iff $\mathfrak{M}, t \models \varphi$ *for all those worlds* $t \sim_a s$ *that are maximal in the order* $\leq_a xy$ *in the* $\sim_a$*-equivalence class of* s.

But the more general notion in our models is that of a *conditional belief*:

$\mathfrak{M}, s \models B_a^{\psi}\varphi$ iff $\mathfrak{M}, t \models \varphi$ *for all those worlds* $t \sim_a s$ *that are maximal for* $\leq_a xy$ *in the set* $\{u \mid s \sim_a u$ *and* $\mathfrak{M}, u \models \psi\}$.[9]

Conditional beliefs generalize absolute beliefs, which are now definable as $B_a^{\top}\varphi$. They *pre-encode* absolute beliefs that we will have *if* we learn certain things. Indeed, the above semantics for $B_a^{\psi}\varphi$ is formally similar to

[7] In particular, ternary world-dependent plausibility relations are found in the semantics of conditional logic: cf. [Lewis, 1973], [Spohn, 1988], models for games: cf. [Stalnaker, 1999], [Benthem, 2014], as well as in recent logical analyses of major paradigms in epistemology: [Holliday, 2012].

[8] Connected orders are equivalent to the 'sphere models' of conditional logic or belief revision theory (cf. [Grove, 1988], [Segerberg, 2001]) – but in these areas, too, a generalization to pre-orders has been proposed: for instance, in the following works: [Burgess, 1984], [Shoham, 1988], and [Veltman, 1985].

[9] These intuitive maximality formulations must be modified in models allowing infinite sequences in the plausibility ordering. Trivialization can then be avoided as follows (cf. the exposition of plausibility semantics in [Girard, 2008]): $\mathfrak{M}, s \models O^{\psi}\varphi$ iff $\forall t \sim s : \exists u :$ $(t \preceq u$ *and* $\mathfrak{M}, u \models \psi$ *and* $\forall v \sim s$: (*if* $u \preceq v$ *and* $\mathfrak{M}, v \models \psi$, *then* $\mathfrak{M}, v \models \varphi$)).

that for conditional assertions $\psi \Rightarrow \varphi$. This allows us to use known results from [Burgess, 1984], [Veltman, 1985]: [10]

Theorem 7 *The logic of $B_a^{\psi}\varphi$ is axiomatized by standard propositional logic plus the laws of conditional logic over pre-orders.*

Deductively stronger modal logics also exist in this area, such as the popular system **KD45** for absolute belief. The structural content of their additional axioms can be determined through standard modal frame correspondence techniques (see [Blackburn *et al.*, 2007], [Benthem, 2010]).

Digression: Further relevant attitudes. Modeling agency with just the notions of knowledge and belief is mainly a tradition inherited from the literature. In a serious study of agency the question needs to be raised afresh what is our natural repertoire of attitudes triggered by information. As one interesting example, the following operator has emerged recently, in between knowledge and belief qua strength. Intuitively, 'safe belief' is belief that agents have which cannot be falsified by receiving true new information.[11] Over epistemic plausibility models $\mathfrak{M}$, its force is as follows:

Definition 8 *The modality of* safe belief $B_a^{+}\varphi$ *is interpreted as follows:*

$$\mathfrak{M}, s \models B_a^{+}\varphi \quad \text{iff} \quad \textit{for all worlds } t \sim_a s\textit{: if } s \leq_a t\textit{, then } \mathfrak{M}, t \models \varphi.$$

Thus, the formula φ is to be true in all accessible worlds that are at least as plausible as the current one. This includes the most plausible worlds, but it need not include all epistemically accessible worlds, since the latter may have some less plausible than the current one. The logic for safe belief is just **S4**, since it is in fact the simplest modality over the plausibility order.

A notion like this has the conceptual advantage of making us see that agents can have more responses to information than just knowledge and belief. [12] But there is also the technical advantage that the simple modality of safe belief can define more complex notions such as conditional belief (see [Lamarre, 1991], [Boutilier, 1994], [Benthem, 2014]) – which can lead to simplifications of logics for agency.

[10]For some recent completeness theorems in deontic logic over Hanson-style betterness orders paralleling this line of work in conditional logic, see our final section on related literature in this chapter.

[11]This notion has been proposed independently in AI, cf. [Shoham and Leyton-Brown, 2008], philosophy, cf. [Stalnaker, 2006], learning theory, and game theory, cf. [Baltag *et al.*, 2011], [Baltag *et al.*, 2009].

[12]Other relevant notions extending the usual epistemic-doxastic core vocabulary include the 'strong belief' of [Stalnaker, 2006], [Baltag and Smets, 2008].

2.4 Dynamic logics of belief change

Having set up the basic attitudes, we now want to deal with explicit acts or events that update not just knowledge, but also agents' beliefs. [13]

Hard information The first obvious triggering event are the earlier public announcements of new hard information. Their complete logic of belief change can be developed in analogy with the earlier dynamic epistemic logic PAL, again via world elimination. Its key recursion axiom for new beliefs uses conditional beliefs:

FACT 1 *The following formula is valid in our semantics:*

$$[!\varphi]B_a\psi \leftrightarrow (\varphi \rightarrow B_a[!\varphi]^{\varphi}\psi)$$

To keep the complete dynamic language in harmony, we then also need a recursion axiom for the conditional beliefs that are essential here:

Theorem 9 *The dynamic logic of conditional belief under public announcements is axiomatized completely by*

(a) any complete static logic for the model class chosen,

(b) the PAL recursion axioms for atomic facts and Boolean operations,

(c) the following recursion axiom for conditional beliefs:

$$[!\varphi]B_a^{\chi}\psi \leftrightarrow (\varphi \rightarrow B_a[!\varphi]^{\varphi\wedge[!\varphi]\chi}\psi)$$

This analysis also extends to the further notion of safe belief, with the following even simpler recursion law:

FACT 2 *The following PAL-style axiom holds for safe belief:*

$$[!\varphi]B_a^{+}\psi \leftrightarrow (\varphi \rightarrow B_a^{+}(\varphi \rightarrow [!\varphi]\psi)).$$

Using this equivalence, which behaves more like the original central PAL axiom, one can show that safe belief has its intuitively intended feature. Safe belief in factual propositions (i.e., those not containing epistemic or doxastic operators) remains safe belief after updates with hard factual information.[14]

[13] For a much more extensive up-to-date treatment of logic-based belief revision, cf. [Benthem and Smets, 2015].

[14] Unlike with plain belief, the latter recursion does not involve a move to an irreducible new notion of 'conditional safe belief'. Indeed, given a definition of conditional belief in terms of safe belief, the more complex recursion law in Theorem 10 can be derived.

Soft information But belief change also involves more interesting triggers, depending on the quality of the incoming information, or the trust agents place in it. 'Soft information upgrade' does not eliminate worlds as what hard information does, but it rather *changes the plausibility order*, promoting or demoting worlds according to their properties. Here is one widely used way in which this order change can happen: an act of 'radical', or 'lexicographic' upgrade.[15]

Definition 10 *A* radical upgrade $\Uparrow\varphi$ *changes the current plausibility order* $\leq$ *between worlds in* $\mathfrak{M}, s$ *to create a new model* $\mathfrak{M}\Uparrow\varphi, s$ *where all* φ*-worlds in* $\mathfrak{M}, s$ *become better than all* $\neg\varphi$*-worlds, while, within those two zones, the old plausibility order* $\leq$ *remains as it was.*

No worlds are eliminated here, it is the ordering pattern that adapts. There is a matching upgrade modality for this in our dynamic language:

$$\mathfrak{M}, s \models [\Uparrow\varphi]\psi \text{ iff } \mathfrak{M}\Uparrow\varphi, s \models \psi.$$

This extended setting supports one more dynamic completeness theorem (cf. [Benthem, 2007]).

Theorem 11 *The logic of radical upgrade is axiomatized completely by*

(a) a complete axiom system for conditional belief on the static models,

(b) the following recursion axioms for postconditions:

$$
\begin{aligned}
&[\Uparrow\varphi]q \leftrightarrow q, \quad \textit{for all atomic proposition letters } q\\
&[\Uparrow\varphi]\neg\psi \leftrightarrow \neg[\Uparrow\varphi]\psi\\
&[\Uparrow\varphi](\psi \wedge \chi) \leftrightarrow ([\Uparrow\varphi]\psi \wedge [\Uparrow\varphi]\chi)\\
&[\Uparrow\varphi]B^{\chi}\psi \leftrightarrow (E(\varphi \wedge [\Uparrow\varphi]\chi) \wedge B([\Uparrow\varphi]^{\varphi \wedge [\Uparrow\varphi]\chi}\psi))\\
&\qquad \vee (\neg E(\varphi \wedge [\Uparrow\varphi]\chi) \wedge B([\Uparrow\varphi]^{[\Uparrow\varphi]\chi}\psi)
\end{aligned}
$$

Here the operator 'E' is the existential epistemic modality, and we need to add a simple recursion axiom for knowledge, that we forego here.[16]

There are many further policies for changing plausibility order. For instance, 'conservative upgrade' $\uparrow\varphi$ only puts the *most plausible* φ-worlds on top in the new model, leaving the rest in their old positions. [Rott, 2006] is an excellent philosophical source for the variety of policies found in belief revision theory that is not tied to the specific dynamic logic methodology employed in this chapter. For general results on complete dynamic logics of belief change in our style, see [Benthem, 2007], [Baltag and Smets, 2008] and

[15]In this section, we drop epistemic accessibility, and focus on plausibility order only.
[16]As before, it is easy to extend this analysis of soft upgrade to the case of safe belief.

[Benthem, 2011]. The most up-to-date survey as of now is the Handbook chapter [Benthem and Smets, 2015].

A plea for patience. Readers wondering why we are introducing all these different notions about information, knowledge and belief, may want to think at this stage already about their counterparts for deontic notions. In fact, analogies are easy to find. For instance, concerning our static repertoire, safe belief is like the 'betterness' modality that we will use later to describe preference. And as for our dynamic repertoire, the distinction between hard and soft information has obvious counterparts in different forces that we can give to commands coming from moral authorities.

2.5 General dynamic methodology and its applications

We have spent quite some time on the above matters because they represent a general methodology of *model transformation* that works for many further phenomena, including changes in preference, and the even richer deontic scenarios that we will be interested in eventually.

Model transformations of relevance to agency can be much more drastic than what we have seen here, extending the domains of available worlds and modifying their relational structure accordingly. In the dynamic epistemic logic of general observation DEL, different agents can have different access to the current informational event, as happens in card games, communication with security restrictions, or other social scenarios. This requires generalizing PAL as well as the above logics of belief change, using a mechanism of 'product update' to create more complex new models (cf. [Baltag *et al.*, 1998], [van Ditmarsch *et al.*, 2007], [Benthem, 2011]).

Appropriately extended update mechanisms have been applied to many further aspects of agency: changes in intentions ([Roy, 2008], [Icard III *et al.*, 2010]), trust ([Holliday, 2009]), inference ([Velazquez-Quesada, 2009]), questions and inquiry ([Benthem and Minica, 2009]), as well as complex scenarios in games ([Otterloo, 2005], [Benthem, 2014]) and social information phenomena generally ([Seligman *et al.*, 2013], [Baltag *et al.*, 2013], [Liu *et al.*, 2014], [Hansen and Hendricks, 2014]). There are also studies tying update mechanisms to general logics of graph change, such as [Aucher *et al.*, 2009a], [Aucher *et al.*, 2016].Yet, in this chapter, we will stick mainly with the much simpler pilot systems presented in the preceding sections.

3 DEONTIC LOGIC AS PREFERENCE LOGIC

Having set up the machinery for changing informational attitudes, we now turn to our major interest in this chapter, the realm of normative evaluation for worlds or actions and the matching dynamic deontic logics. Here we will follow a perhaps not uncontroversial track: our treatment of deontic notions and scenarios will be based on *preference* structure and its changes. We believe that this is a conceptually good way of looking at deontic notions, and at the same time, it lends itself very well to treatment by our earlier methods, since at an abstract level, doxastic plausibility order and deontic betterness order are very similar. [17] The results that follow in the coming sections are largely from [Liu, 2008], [Girard, 2008], and [Liu, 2011a].[18]

Let us say a few more words about the connection between deontic logic and preference, to justify our approach in this chapter. Deontic logic is the logical study of normative concepts such as obligation, prohibition, permission and commitment. This area was initiated by von Wright in [von Wright, 1951] who introduced the logic of absolute obligation. As a reaction to paradoxes with this notion, conditional obligation was then proposed in [von Wright, 1956], [von Wright, 1964] and [Fraassen, 1972]. Good reviews systematizing the area are found in [Åqvist, 1987], [Åqvist, 1994].

One often thinks of deontic logic as the study of some accessibility relation from the actual world to the set of 'ideal worlds', but the more sophisticated view ([Hansson, 1969], [Fraassen, 1973] and [Jackson, 1985]) has models with a binary comparison relation.[19] Such more general comparisons between worlds make sense, for instance, when talking and reasoning about 'the lesser of two evils', or about 'improvement' of some given situation.

This is precisely the ordering semantics we already saw for belief, and it would be tedious to indulge in formal definitions at this stage that the reader can easily construct for herself. Our base view is that of binary *pre-orders* as before, for which we will now use the notation R to signal a change from the earlier plausibility interpretation. As usual, imposing further constraints on the ordering will generate deductively stronger deontic logics. The binary relation R now interprets $O\varphi$ (absolute obligation) as φ *being true in all best worlds*, much like belief with respect to plausibility. Then conditional obligation $O^{\psi}\varphi$ is like conditional belief: φ holds *in the best ψ-worlds*. [20]

For further information on deontic logic, we refer to [Åqvist, 1994] and various chapters in this Handbook [Gabbay *et al.*, 2013]. Our emphasis in this chapter will be mainly on interfacing with this field.

---

[17]The stated parallel is of course also well-known from the deontic literature, for instance, in the works of Hansson or van der Torre cited in our text.

[18]To unclutter notation, here and henceforth, we will mostly suppress agent indices for modal operators in our languages and their corresponding semantic relations. While we believe that deontic scenarios are very often essentially multi-agent in nature, it is useful to stay with single agent notations as long as these suffice.

[19]Hansson argued that von Wright-type deontic logic can be naturally interpreted in terms of a preference relation 'is at least as ideal as' among possible worlds – an ordering that we will call 'betterness' in what follows. This research program in deontic logic is still very much alive today, witness the chapter by Xavier Parent in this Handbook.

[20]There are also more abstract neighborhood versions of this semantics, where the current proposition plays a larger role in terms of binary deontic betterness relations R^{ψ}, where one can set $\mathfrak{M}, s \models O^{\psi}\varphi$ iff for all t in W with $sR^{\psi}t, \mathfrak{M}, t \models \varphi$.

As we already noted at the start of this paper, deontic ordering shows intuitive analogies with the notion of *preference*. One can think of betterness as reflecting the preferences of a moral authority or law-giver, and in the happy Kantian case where agents' duties coincide with their inclinations, deontic betterness *is* in fact the agent's own preference. We claim no novelty for this line of thought, which was advocated forcefully as early as [Hansson, 1969]. With this twist, we can then avail ourselves of existing studies of preference structure and evaluation dynamics, a line of thinking initiated in [van der Torre, 1997] and [van der Torre and Tan, 1999], though we now take the dynamic-epistemic road.

By way of background to what follows, we note that preference logic is a vigorous subject with its own history. For many new ideas and results in the area, we refer to [Hansson, 2001a] and [Grune-Yanoff and Hansson, 2009], while our final section on related literature is also relevant. What we will do next in this chapter is discuss some major recent developments in the study of preference statics and dynamics, emphasizing those that we see as being of relevance to deontic logic, an area where we will return explicitly later on in this chapter. [21]

4 STATIC PREFERENCE LOGIC

In the coming sections, we will discuss basic developments in modal preference logic, starting with its statics, and continuing with dynamics of preference change. Our treatment follows pioneering ideas from [Boutilier, 1994] and [Halpern, 1997], and for the dynamics, we mainly rely on [Benthem *et al.*, 2006] and [Benthem and Liu, 2007].

4.1 General modal preference logic

Our basic models are like in decision theory or game theory: there is a set of alternatives (worlds, outcomes, objects) ordered by a primitive ordering that we dub 'betterness' to distinguish it from richer notions of preference.[22]

Definition 12 *A* modal betterness model *is a tuple $\mathfrak{M} = (W, \preceq, V)$ with W a set of worlds or objects, $\preceq$ a reflexive and transitive relation over these, and V is a valuation assigning truth values to proposition letters at worlds.*[23]

[21] Preference logic tends to focus on the agents' own preferences, not those of others, but it applies equally well to multi-agent settings such as social choice problems, decisions in games, or moral scenarios, where different preference orders interact in crucial ways.

[22] To repeat an earlier point, while each agent has her own betterness order, in what follows, merely for technical convenience, we suppress indices wherever we can.

[23] As we said before, we use pre-orders since we want the generality of possibly non-total preferences. Still, total orders, the norm in areas like game theory, provide an interesting specialization for the results in this chapter – but we will only mention it in passing.

The order relation in these models also induces a strict variant $s \prec t$:

If $s \preceq t$ but not $t \preceq s$, then t is *strictly better* than s.

Here is a simple modal language that can say a lot about these structures:

Definition 13 *Take any set of propositional variables Φ, with p ranging over Φ. The* modal betterness language *has this inductive syntax rule:*

$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle\leq\rangle\varphi \mid \langle<\rangle\varphi \mid E\varphi.$

The intended reading of $\langle\leq\rangle\varphi$ is "φ is true in a world that is at least as good as the current world", while $\langle<\rangle\varphi$ says that "φ is true in a world that is strictly better than the current world." In addition, the auxiliary *existential modality* $E\varphi$ says that "there is a world where φ is true". Also, as usual, we write $[\leq]\varphi$ for the defined universal modality $\neg\langle\leq\rangle\neg\varphi$, and we use $[<]$ and U for the duals of $\langle<\rangle\varphi$ and E, respectively. Combinations of these modalities can capture a wide variety of binary preference statements comparing propositions, witness the cited literature.

The interpretation of this modal language over our models is standard:

Definition 14 *Truth conditions for the atomic propositions and Boolean combinations are standard. Modalities are interpreted like this:*

- $\mathfrak{M}, s \models \langle\leq\rangle\varphi$ *iff* *for some t wih $s \preceq t$, $\mathfrak{M}, t \models \varphi$.*
- $\mathfrak{M}, s \models \langle<\rangle\varphi$ *iff* *for some t with $s \prec t$, $\mathfrak{M}, t \models \varphi$.*
- $\mathfrak{M}, s \models E\varphi$ *iff* *for some world t in W, $\mathfrak{M}, t \models \varphi$.*

The defined modalities use the obvious universal versions of these clauses. For concreteness, we state the standard calculus to come out of this.

Theorem 15 *Modal betterness logic is completely axiomatized by*

1. *the system* ***S4*** *for the preference modality,*
2. *the system* ***S5*** *for the universal modality,*
3. *the connecting law $U\varphi \to [\preceq]\varphi$,*
4. *three axioms that govern the strict betterness modality and its interaction with the weak preference modality: cf.[Benthem* et al.*, 2009c].*

4.2 Special features of preference

Next we briefly survey three special logical features of preference structure that go beyond standard modal logic of pre-orders, and that will eventually turn out to be of interest to deontics as well.

Lifting to generic preferences. While betterness relates specific objects or worlds, preference is often used generically for comparing different *kinds* of things. Ever since [von Wright, 1963], logicians have also studied preferences $P(\varphi, \psi)$ between propositions, viewed as properties of worlds, or of objects.

There is not one such notion, but many, that can be defined by a *lift* of the betterness order among worlds to sets of worlds, cf. [Halpern, 1997], [Benthem *et al.*, 2009c], [Liu, 2011a]. For instance, compare your next moves in a game, identified with the set of outcomes that they lead to. Which move is 'better' depends on the criterion chosen: maybe we want to go with the one leading to the highest possible outcome, or the one with the highest minimally guaranteed outcome, etcetera.

Such options are reflected in various quantifier combinations for the lifting. In particular, von Wright had a $\forall\forall$-type preference between sets P, Q:

$$\forall x \in P\ \forall y \in Q\colon\ x \preceq y.$$

A simpler also useful example is the modal $\forall\exists$-type

$$\forall x \in P\ \exists y \in Q\colon\ x \preceq y.$$

This says that for any P-world, there is a Q-world which is at least as good as that ψ-world. In the earlier-mentioned game setting, this stipulation would say that the most preferred moves have the highest maximal outcomes. Unlike the $\forall\forall$-version, this ubiquitous $\forall\exists$ generic preference can be defined in the above modal preference language, using the universal modality ranging over all worlds:

$$P^{\forall\exists}(\varphi, \psi) := U(\psi \to \langle\leq\rangle\varphi).$$

This generic preference $P\varphi\psi$ satisfies the usual properties for preference, reflexivity and transitivity: for instance, $P\varphi\psi$ and $P\psi\chi$ imply $P\varphi\chi$.[24]

Ceteris paribus clauses. Unlike plausibility, preference ordering seldom comes in pure form: the comparison between alternatives is often entangled with other considerations. Again, games provide an example. Usually, players do not compare moves via the sets of all their possible outcomes, but rather, they compare the *most plausible* outcomes of their moves. This is the so-called *normality sense* of ceteris paribus preference: we do not compare all the φ and ψ-worlds, but only the 'normal ones' in some relevant

[24]Other quantifed stipulations lead to other generic preferences. This proliferation may be a problem (e.g., 'doing what is best' then depends on one's stipulation as to what 'best' means), but there is no consensus in the literature that one can appeal to. A logical approach at least helps make the options clear.

sense. This belief restriction, observed by many authors, will return in our discussion of doxastic entanglement of preference in Section 8.

But there are also other natural senses of taking a ceteris paribus clause. It was noticed already in [von Wright, 1963] that there is also an 'equality sense' of preference, involving a hidden assumption of *independence*. In that case, one only make comparisons between worlds where some things or issues are held constant, in terms of giving the same truth values to some specified set of atomic propositions, or complex formulas. The logic of equality-based preference is of independent interest, and it has been axiomatized and analyzed in detail in [Benthem *et al.*, 2009c].

Richer preference languages. Modal languages are just one step on a ladder of formalisms for analyzing reasoning practices. It has been claimed that richer languages are needed to faithfully render basic preference notions, cf. [de Jongh and Liu, 2009] on first-order preferences among objects, [Grandi and Endriss, 2009] on first-order languages of social choice, [Benthem *et al.*, 2006] on hybrid modal preference languages for defining backward induction solutions in games, the hybrid modal language of 'desire' and 'freedom' for decision making in [Guo and Sliegman, 2011], or the modal fixed-point languages for games used in [Benthem, 2014]. Though we will mainly use modal formalisms to make the essential points of this paper, we will mention the relevance of such richer preference formalisms occasionally.

5 WORLD BASED DYNAMICS OF PREFERENCE CHANGE

Now let us look at how given preferences can change. Intuitively, there are many acts and events that can have such an effect. Perhaps the purest form is a radical *command* by some moral authority to do something. This makes the worlds where we act better than those where we do not (cf. [Yamada, 2006], a pioneering study on the dynamics of deontic commands): at least, if we 'take' the order as a legitimate instruction, and change our evaluation accordingly, overriding any preferences that we ourselves might have had. Technically, this dynamics will change a current betterness relation in a model. These phenomena can be studied entirely along the lines already developed here for information dynamics.[25]

5.1 Betterness change

[Benthem and Liu, 2007] is a first systematic study of betterness change using methods from dynamic-epistemic logic. The running example in their approach is a weak 'suggestion' $\sharp\varphi$ that a proposition φ be the case. This relatively modest ordering change leaves the set of worlds the same, but it

[25] Of earlier treatments, we mention [van der Torre and Tan, 1999], based on the dynamic semantics for natural language proposed in [Veltman, 1996].

removes any preferences that the agent might have had for $\neg\varphi$-worlds over φ-worlds among these.[26]

Caveat We are not claiming that the technical notion of suggestion as defined here is the most basic action of preference change or deontic change. We start with this system merely as a simple pilot version for our methodology, just as we did with PAL for information change.

The main general point to note here is that events with evaluative import can act as triggers that change some current betterness relation on worlds. In particular, a suggestion $\sharp\varphi$ leads to the following model change:

Definition 16 *Given any modal preference model* $(\mathfrak{M}, s)$*, the* suggestion upgrade $(\mathfrak{M}\sharp\varphi, s)$ *has the same domain, valuation, and actual world as* $(\mathfrak{M}, s)$*, but the new preference relations are now*

$$\preceq_i^* = \preceq_i - \{(s,t) \mid \mathfrak{M}, s \models \varphi \textit{ and } \mathfrak{M}, t \models \neg\varphi\}$$

In preference models $\mathfrak{M}$, a matching dynamic modality is interpreted as:

$$(\mathfrak{M}, s) \models [\sharp\varphi]\psi \quad \text{iff} \quad \mathfrak{M}_{\sharp\varphi}, s \models \psi$$

Again, complete dynamic logics exist (cf. [Benthem and Liu, 2007]). The reader will find it particularly useful to scrutinize the key recursion law for preferences after suggestion.[27]

Theorem 17 *The dynamic preference logic of suggestion is completely axiomatized, over its static base logic, by the following principles:*

1. $\langle\sharp\varphi\rangle p \;\leftrightarrow\; p$
2. $\langle\sharp\varphi\rangle\neg\psi \;\leftrightarrow\; \neg\langle\sharp\varphi\rangle\psi$
3. $\langle\sharp\varphi\rangle(\psi\wedge\chi) \;\leftrightarrow\; (\langle\sharp\varphi\rangle\psi \wedge \langle\sharp\varphi\rangle\chi)$
4. $\langle\sharp\varphi\rangle\langle\leq\rangle\psi \;\leftrightarrow\; (\neg\varphi \wedge \langle\leq\rangle\langle\sharp\varphi\rangle\psi) \vee (\langle\leq\rangle(\varphi \wedge \langle\sharp\varphi\rangle\psi))$
5. $\langle\sharp\varphi\rangle E\psi \;\leftrightarrow\; E\langle\sharp\varphi\rangle\psi$

Similar completeness results are presented in [Liu, 2011a] for dynamic logics that govern many other kinds of normative action, such as the 'strong commands' corresponding to our earlier radical plausibility upgrade. Following the latter instruction, deontically, the agent incorporates the wish of some over-riding authority.

[26] Similar operations have come up recently in logical treatments of relevant alternatives theories in epistemology, when modeling changes in what is considered relevant to making or evaluating a knowledge claim. Cf. [Holliday, 2014], [Benthem, 2016a].

[27] Technically, the simplicity of this law reflects the clear analogy between our universal preference modality and the earlier doxastic notion of safe belief.

Deontic logicians (or linguists interested in speech acts) will find it easy to come up with many further normative triggers in between weak suggestions and strong commands, but the above-mentioned methods can deal with a wide variety of such proposals.

5.2 Deriving changes in defined preferences

This is an analysis of betterness change and modal statements about it local to specific worlds. But it also applies to the earlier lifted *generic preferences.* As an illustration, consider the $\forall\exists$-lift defined earlier:

FACT 3 *The following equivalence holds for generic* $\forall\exists$ *preference:*

$$\langle\sharp A\rangle P^{\forall\exists}(\varphi,\psi) \quad \textit{iff} \quad P^{\forall\exists}(\langle\sharp A\rangle\varphi, \langle\sharp A\rangle\psi) \wedge P^{\forall\exists}((\langle\sharp A\rangle\varphi \wedge A), (\langle\sharp A\rangle\psi \wedge A)).$$

We omit the simple calculation for this outcome. Similar results may be obtained for other set liftings such as Von Wright's $\forall\forall$-version.

Finally, the recursive style of dynamic analysis presented here also applies to various forms of ceteris paribus preference; cf. [Girard, 2008].

5.3 General formats for betterness change

Behind our specific examples of betterness change, there lies a much more general theory that works for a wide class of triggering events that change betterness or evaluation order. One widely applicable way of achieving greater generality uses programs from *propositional dynamic logic PDL.*

For instance, suggesting that φ is defined by the program:

$$\sharp\varphi(R) := (?\varphi; R; ?\varphi) \cup (?\neg\varphi; R; ?\neg\varphi) \cup (?\neg\varphi; R; ?\varphi).$$

where R is the given input relation, while the operations $?\varphi$ test whether the relevant proposition φ, or related ones, hold. In particular, the disjunct $(?\varphi; R; ?\varphi)$ means that we keep all old betterness links that run from φ-worlds to φ-worlds.

The preceding definition is equivalent in the dynamic logic PDL to the following more compact program expression

$$\sharp\varphi(R) := (?\neg\varphi; R) \cup (R; ?\varphi).$$

Again this keeps all old R-links as they were, except for deleting those that ran from φ-worlds to $\neg\varphi$-worlds.

Likewise, our plausibility changers for belief revision can be defined in this format. For instance, the earlier 'radical upgrade' is defined by

$$\Uparrow\varphi(R) := (?\varphi; R; ?\varphi) \cup (?\neg\varphi; R; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi)$$

Here the constant symbol $\top$ denotes the universal relation that holds between any two worlds. This reflects the original meaning of this transformation: all φ-worlds become better than all $\neg\varphi$-worlds, whether or not they were better before, and within these two zones, the old ordering remains.[28]

Given any PDL program definition of the above sort, one can automatically write recursion laws for the complete dynamic logic of its induced model change, cf. [Benthem and Liu, 2007] for the precise algorithm that computes these axioms. As an illustration, here is the straightforward computation for suggestions:

$$\begin{aligned}
\langle \sharp\varphi \rangle \langle R \rangle \psi &\leftrightarrow \langle (?\neg\varphi; R) \cup (R; ?\varphi) \rangle \langle \sharp\varphi \rangle \psi \\
&\leftrightarrow \langle ?\neg\varphi; R \rangle \langle \sharp\varphi \rangle \psi \vee \langle R; ?\varphi \rangle \langle \sharp\varphi \rangle \psi \\
&\leftrightarrow (\neg\varphi \wedge \langle R \rangle \langle \sharp\varphi \rangle \psi) \vee \langle R \rangle (\varphi \wedge \langle \sharp\varphi \rangle \psi).
\end{aligned}$$

For alternative general formats of ordering change supporting our sort of dynamic logics, we refer to the 'priority update' with event models proposed in [Baltag and Smets, 2008], the general order merge perspective developed in [Benthem, 2006], as well as the still more general 'dynamic dynamic logic' of [Girard *et al.*, 2012].

In our view, the practical and theoretical theoretical variety of ordering changes for plausibility and preference is not a nuisance, but a feature. It matches the wealth of evaluative actions that we encounter in daily life.

6 REASON-BASED PREFERENCES

Primitive betterness relations among worlds or objects reflect what are called 'intrinsic preferences'. But very often, our preferences have underlying structure, and we compare according to criteria: our preferences are then reason-based, or 'extrinsic'. In this section we develop the latter view, that has motivations in linguistic Optimality Theory, cf. [Prince and Smolensky, 2004], and belief revision based on entrenchment, cf. [Rott, 2003]. This view also occurs in reason-based deontic logic, cf. [Fraassen, 1973], probably the first paper ever to propose the style of thinking in this chapter, [Goble, 2000] and [Jackson, 1985], as we shall see in Section 9.

A simple illustration of our approach, that suffices for many natural scenarios, starts with the special case of linear orders for relevant properties that serve as criteria for determining our evaluation of the comparative merits of objects or worlds.

[28] Conservative upgrades can be dealt with in a similar way. As commands, these leave the agent more of her original preferences: so, differences with radical commands will show up in judgments of 'conditional betterness', as discussed in the literature on conditional obligation: see [Hansson, 1969].

6.1 Priority based preference

The following proposal has many ancestors, among which we mention the treatment in [Friedman and Halpern, 1995],[Rott, 2003]. We follow [de Jongh and Liu, 2009], that starts from a given primitive ordering among propositions ('priorities' among properties of objects or worlds), and then derives a preference among objects themselves.

Definition 18 *A* priority sequence *is a finite linear sequence of formulas written as follows:* $C_1 \gg C_2 \cdots \gg C_n \quad (n \in \mathbb{N})$*, where the* C_m *come from a language describing objects, with one free variable* x *in each* C_m*.*

Definition 19 *Given a priority sequence and objects* x *and* y*, Pref(x, y) is defined lexicographically: at the first property* C_i *in the given sequence where* x, y *have a different truth value,* $C_i(x)$ *holds, but* $C_i(y)$ *fails.*

The logic of this framework is analyzed in [de Jongh and Liu, 2009], while applications to deontic logic are developed in [Benthem *et al.*, 2010].

As it happens, this is only one of many ways of deriving a preference ordering from a given priority sequence. A good overview of existing approaches is found in [Coste-Marquis *et al.*, 2004].

6.2 Pre-orders

In general, comparison orders need not be connected, and then the preceding needs a significant generalization. This was done, in a setting of social choice and belief merge, in the seminal paper [Andréka *et al.*, 2002], which we adapt slightly here to the notion of 'priority graphs', based on the treatment in [Girard, 2008], [Liu, 2011b].

The following definitions contain a free parameter for a *language* L that can be interpreted in the earlier modal betterness models $\mathfrak{M}$. For simplicity only, we will take this to be a simple propositional language of properties.

Definition 20 *A* priority graph $\mathscr{G} = \langle P, < \rangle$ *is a strictly partially ordered set of propositions in the relevant language of properties* L*.*

Here is how one derives a betterness order from a priority graph:

Definition 21 *Let* $\mathscr{G} = \langle P, < \rangle$ *be a priority graph, and* $\mathfrak{M}$ *a model in which the language* L *defines properties of objects. The* induced betternness relation $\preceq_{\mathscr{G}}$ *between objects or worlds is defined as follows:*

$$y \preceq_{\mathscr{G}} x := \forall P {\in} \mathscr{G}((Py \to Px) \vee \exists P' {<} P(P'x \wedge \neg P'y)).$$

Here, in principle, $y \preceq_{\mathscr{G}} x$ requires that x has every property in the graph that y has. But there is a possibility of 'compensation': if y has P while x does not, this is admissible, provided there is some property P' with higher priority in the graph where x does better: x has P' while y lacks it. Clearly, this stipulation subsumes the earlier priority sequences: linear priority graphs lead to lexicographic order.

One can think of priority graphs of propositions in many ways that are relevant to this paper. In the informational realm, they are hierarchically ordered information sources, structuring the evidence for agents' beliefs. In the normative realm, they can stand for complex hierarchies of laws, or of norm givers with relative authority.

6.3 Static logic and representation theorem

In what follows, we immediately state a crucial technical property of this framework, cf. [Friedman and Halpern, 1995], [Liu, 2011b].

Theorem 22 *Let $\mathfrak{M} = (W, \preceq, V)$ be any modal preference model, without constraints on its relation. The following two statements are equivalent:*

(*a*) *The relation $y \preceq x$ is a reflexive and transitive order,*

(*b*) *There is a priority graph $\mathscr{G} = (P, <)$ such that,*
for all worlds $x, y \in W$, $y \preceq x$ iff $y \preceq_{\mathscr{G}} x$.

This representation theorem says that the general logic of derived extrinsic betterness orderings is still just that of pre-orders. But it also tells us that any intrinsic pre-order can be rationalized as an extrinsic reason-based one by adding structure without disturbing the base model as it is.

6.4 Priority dynamics and graph algebra

Now, we have a new locus for more fine-grained preference change: the family of underlying reasons, which brings its own logical structure. For linear priority sequences, relevant changes involve the obvious operations $[^+C]$ of adding a new proposition C to the right, $[C^+]$ of adding C to the left, and various functions $[-]$ dropping first, last or intermediate elements of a priority sequence. [de Jongh and Liu, 2009] give complete dynamic logics for these. Here is one typical valid principe:

$$[^+C]Pref(x,y) \leftrightarrow Pref(x,y) \vee (Eq(x,y) \wedge C(x) \wedge \neg C(y))$$

Operations for changing preferences become even richer in the realm of priority graphs, due to their possibly non-linear structure. However, in this setting an elegant mathematical alternative arises, in terms of an algebra of merely two fundamental operations that combine arbitrary graphs:

- $\mathscr{G}_1 ; \mathscr{G}_2$ adding a graph to another in top position

- $\mathscr{G}_1 \| \mathscr{G}_2$ adding two graphs in parallel.

One can think of this as the obvious graph operations of 'sequential' and 'parallel' composition. Here the very special case where one of the graphs consists of just one proposition models our earlier simple update actions.

This graph calculus has been axiomatized completely in [Andréka *et al.*, 2002] by algebraic means, while [Girard, 2008] presents a further modal-style axiomatization. We display its major modal principles here, since they express the essential recursion underlying priority graph dynamics.

Here is one case where, as mentioned earlier, a slight language extension is helpful: in what follows, the proposition letter n is a 'nominal' from hybrid logic denoting one single world.

$$\langle \mathscr{G}_1 \| \mathscr{G}_2 \rangle^{\leq} n \leftrightarrow \langle \mathscr{G}_1 \rangle^{\leq} n \wedge \langle \mathscr{G}_2 \rangle^{\leq} n.$$

$$\langle \mathscr{G}_1 \| \mathscr{G}_2 \rangle^{<} n \leftrightarrow (\langle \mathscr{G}_1 \rangle^{<} n \wedge \langle \mathscr{G}_2 \rangle^{\leq} n) \vee (\langle \mathscr{G}_1 \rangle^{\leq} n \wedge \langle \mathscr{G}_2 \rangle^{<} n).$$

$$\langle \mathscr{G}_1 ; \mathscr{G}_2 \rangle^{\leq} n \leftrightarrow (\langle \mathscr{G}_1 \rangle^{\leq} n \wedge \langle \mathscr{G}_2 \rangle^{\leq} n) \vee \langle \mathscr{G}_1 \rangle^{<} n.$$

$$\langle \mathscr{G}_1 ; \mathscr{G}_2 \rangle^{<} n \leftrightarrow (\langle \mathscr{G}_1 \rangle^{\leq} n \wedge \langle \mathscr{G}_2 \rangle^{<} n) \vee \langle \mathscr{G}_1 \rangle^{<} n.$$

These axioms reduce complex priority relations to simple ones, after which the whole language reduces to the modal logic of weak and strict atomic betterness orders. In particular, this modal graph logic is decidable.

Thus, we have shown how putting reasons underneath agents' preferences (or, for that matter, their beliefs) admits of precise logical treatment, while still supporting the systematic dynamics that we are after.

7 A TWO-LEVEL VIEW OF PREFERENCE

Now we have two ways of looking at preference: one through intrinsic betterness order on modal models, the other through priority structure giving reasons inducing extrinsic betterness orders. One might see this as calling for a reduction from one level to another, but instead, *combining* the two perspectives seems the more attractive option, as providing a richer modeling tool for preference-driven agency.

7.1 Harmony of world order and reasons

In many cases, the two modeling levels are in close harmony, allowing for easy switches from one to the other (cf. [Liu, 2008]):

Definition 23 *Let α: $(\mathscr{G}, A) \to \mathscr{G}'$, with $\mathscr{G}$, $\mathscr{G}'$ priority graphs, and A a new proposition. Let σ be a map from $(\preceq, A)$ to $\preceq'$, where $\preceq$ and $\preceq'$ are betterness relations over worlds. We say that α* induces σ*, if always:*

$$\sigma(\preceq_{\mathscr{G}}, A) \quad = \quad \preceq_{\alpha(\mathscr{G},A)}$$

Here are two results that elaborate the resulting harmony between two levels for our earlier major betterness transformers:

FACT 4 *Taking a suggestion A is the map induced by the priority graph update $\mathscr{G}\|A$. More precisely, the following diagram commutes:*

$$\begin{tikzcd}
\langle \mathscr{G}, < \rangle \arrow[r, "\|A"] \arrow[d] & \langle (\mathscr{G}\|A), < \rangle \arrow[d] \\
\langle W, \preceq \rangle \arrow[r, "\sharp A"] & \langle W, \sharp A(\preceq) \rangle
\end{tikzcd}$$

For a second telling illustration of such harmony in terms of our earlier themes, consider a priority graph $(\mathscr{G}, <)$ with a new proposition A added on top. The logical dynamics at the two levels is now correlated as follows:

FACT 5 *Placing a new proposition A on top of a priority graph $(\mathscr{G}, <)$ induces the radical upgrade operation $\Uparrow A$ on possible worlds ordering models. More precisely, the following diagram commutes:*

$$\begin{tikzcd}
\langle \mathscr{G}, < \rangle \arrow[r, "A;\mathscr{G}"] \arrow[d] & \langle (A;\mathscr{G}), < \rangle \arrow[d] \\
\langle W, \preceq \rangle \arrow[r, "\Uparrow A"] & \langle W, \Uparrow A(\preceq) \rangle
\end{tikzcd}$$

Thus the two kinds of preference dynamics, living at different levels of detail in representing scenarios, dovetail well: [Liu, 2011a] has details.

7.2 Correlated dynamics

There are several advantages to working at both levels without assuming automatic reductions. For a start, not all natural operations on graphs have matching betterness transformers at all. An example from [Liu, 2011b] is *deletion* of the topmost elements from a given priority graph. This syntactic operation of removing criteria is not invariant for replacing graph arguments by other graphs inducing the same betterness order, and hence it is a genuine extension of preference change.

But also conversely, there is no general match. Not all PDL-definable betterness changers from Section 5.3 are graph-definable. In particular, not all PDL transformers preserve the basic order properties of reflexivity and transitivity guaranteed by priority graphs. For a concrete illustration, consider the program

$?A; R$: 'keep the old relation only from where A is true'.

This change does not preserve reflexivity of an order relation R, because the $\neg A$-worlds now have no outgoing relation arrows any more. [29]

All this argues for a more general policy of co-existence, modeling both intrinsic and extrinsic preference for agents, with reasons for the latter explicitly encoded in priority graphs as an explicit part of the modeling. [30]

Coda: Switching perspectives on preference. One might still have a favorite, and think, for instance, that intrinsic betterness relations merely reflect an agent's raw feelings or prejudices. But the intrinsic-extrinsic contrast is relative, not absolute. If I obey the command of a higher moral authority, I may acquire an extrinsic preference, whose reason is the duty of obeying a superior. But for that higher agent, the same preference may well be intrinsic: "The king's whim is my law". This observation suggests a further theme: namely, transitioning from one perspective to the other. We conclude with a few remarks on realizing this option.

7.3 Additional dynamics: language change

Technically, intrinsic betterness can become extrinsic through a dynamics that has been largely outside the scope of dynamic-epistemic logic so far, that of *language change*. One mechanism here is the proof of the earlier representation result stated in Theorem 22. It partitions the given betterness pre-order into clusters, and if these are viewed as new relevant reasons or criteria, the resulting strict order of clusters is a priority graph inducing the given order. This may look like mere formal rationalization, but in practice, one often observes agents' preferences between objects, and then postulates reasons for them. A relevant source is the notion of 'revealed preference' from the economics literature: cf. [Houser and Kurzban, 2002].

Thus, our richer view of preference also suggests a new kind of dynamics beyond what w shave considered so far. In general, reasons for given preferences may have to come from some other, richer language than the one that we started with: we are witnessing a dynamic act of *language creation*.[31]

[29] Intuitively, the operation $?A; R$ amounts to a refusal to make betterness comparisons at worlds that lack property A. Though somewhat idiosyncratic, this seems a bona fide mind change for an agent.

[30] The observations in this section fit well with a general theme in logics of agency today: that of *tracking* dynamic updates operating at finer levels by operations at coarser levels of representing information [Benthem, 2016b]. Tracking is sometimes possible, sometimes it is not, and there are systematic reasons for these phenomena. We will return to this theme in our section on Further Directions.

[31] For a study of language change in the setting for belief revision, cf. [Parikh, 1999].

8 COMBINING EVALUATION AND INFORMATION

We have now completed our exposition of information dynamics as well as preference dynamics, which brought its own further topics. What must have become abundantly clear is that there are strong formal similarities in the logic of order and order change in the two realms. We have not even enumerated all of these similarities, but, for instance, all of our earlier ideas and results about reason-based preference also make sense when analyzing evidence-based belief.

This compatibility helps with the next natural step we must take. As we said right at the start of this paper, the major agency systems of information and evaluation do not live in isolation: they interact all the time. A rational agent can process information well in the sense of proof or observation, but is also 'reasonable' in a broader sense of being guided by goals.

This *entanglement* of knowledge, belief, and preference is essential to how preference functions, [32] and it shows in many specific settings. We will look at a few cases, and in particular, their impact on the dynamics of preference change. [33] This is where we need a combination of all ideas presented so far: static epistemic, doxastic logic, preference logic and deontic logic, as well as dynamic logics of update actions appropriate to all these notions.

Though we will mainly discuss here how information dynamics influences preference and deontic notions, the opposite influence is equally real. In particular, successful information flow depends on *trust* and *authority*: both clearly deontic notions.[34]

8.1 Generic preference with knowledge

In Section 4.2, we defined one basic generic preference as follows:

$$Pref^{\forall\exists}(\psi,\varphi) := U(\psi \to \langle\leq\rangle\varphi).$$

This refers to possibilities in the whole model, including even those that an agent might know to be excluded. [Benthem and Liu, 2007] defend this scenario in terms of 'regret', but still, there is also a reasonable intuition that preference only runs among situations that are epistemically possible.

This suggests the entangled notion that, for any ψ-world that is *epistemically accessible* to agent a in the model, there is a world which is at least as good where φ is true. This can be written with an epistemic modality:

$$Pref^{\forall\exists}(\varphi,\psi) ::= K_a(\psi \to \langle\leq\rangle\varphi). \quad (K_{bett})$$

[32] Think of the crucial notion of *expected value* in making decisions which mixes preference and probability as subjective belief. Less quantitative examples will follow below.

[33] For a more general discussion of deontic-epistemic entanglement, we refer to [Pacuit *et al.*, 2006], to which we will return later in this chapter.

[34] Following Wittgenstein,[Brandom, 1994] has even argued that language use can only be fully understood in terms of commitments that carry rights and obligations.

But this entangled notion is not yet what we are after, since we want the 'better world' to be epistemically accessible itself. [Liu, 2009a] shows how this cannot be defined in a simple combined language of knowledge and betterness, and that instead, a richer preference formalism is needed with a new *intersection modality* for epistemic accessibility and betterness. The latter entangled notion can be axiomatized, and it also supports a dynamic logic of preference change as before.[35]

8.2 Generic preference with belief

Issues of entanglement become even more appealing with generic preference and belief, where the two relational styles of modeling were very similar to begin with. Again, we might start with a mere combination formula

$$Pref^{\forall\exists}(\varphi,\psi) ::= B_a(\psi \to \langle\leq\rangle\varphi). \qquad (B_{bett})$$

This says that, among the most plausible worlds for the agent, for any ψ-world, there exists a world which is at least as good where φ is true.[36]

Again, this seems not quite right in many cases, since we often want the better worlds relevant to preference to stay inside the most plausible part of the model, being 'informational realists' in our desires, not wanting the impossible. To express this, we again need a stronger merge of the two relations by intersection. The key clause for a corresponding new modality then reads like a 'wishful safe belief':

$$\mathfrak{M}, s \models H\varphi \text{ iff } \textit{for all } t \textit{ with both } s \leq t \textit{ and } s \preceq t,\ \mathfrak{M}, t \models \varphi.$$

As before, the static and dynamic logic of this entangled notion yield to the general dynamic-epistemic methodology explained in earlier sections.

8.3 Other entanglements of preference and normality

Entangled versions of plausibility and betterness abound in the literature. For instance, [Boutilier, 1994] has models $\mathfrak{M} = (W, \leq_P, \leq_N, V)$ with W a set of possible worlds, V a valuation function and $\leq_P$, $\leq_N$ two transitive connected relations $x \leq_P y$ ('y is as good as x) and $x \leq_N y$ ('y is as normal as x). He then defines an operator of *conditional ideal goal* (IG):

$$\mathfrak{M} \models IG^{\psi}\varphi \ \text{ iff } \ Max(\leq_P, Max(\leq_N, Mod(\psi))) \subseteq Mod(\varphi)$$

[35] An alternative approach would impose *additional modal axioms* that require betterness alternatives to be epistemic alternatives via frame correspondence. However, this style of working puts constraints on our dynamic operations on models that we have not yet investigated systematically.

[36] One might also think here of using a *conditional belief* $B^{\psi}\langle\leq\rangle\varphi$, but to us, the latter logical form seems to express an intuitively less plausible form of entanglement.

This says that the best of the most normal ψ worlds satisfy φ. Such entangled notions are still expressible in the modal systems of this chapter.

FACT 6 $IG^{\psi}\varphi ::= (\psi \wedge \neg\langle B^{<}\rangle\psi) \wedge \neg\langle < \rangle(\psi \wedge \neg\langle B^{<}\rangle\psi) \rightarrow \varphi$[37]

Following up on this, now more in the tradition in agency studies in computer science, [Lang *et al.*, 2003] defines the following normality-entangled notion of preference:

Definition 24 *$\mathfrak{M} \models Pref^*(\varphi, \psi)$ iff for all $w' \in Max(\leq_N, Mod(\psi))$ there exists $w \in Max(\leq_N, Mod(\varphi))$ such that $w' <_P w$.*

This reflects one of the earlier-mentioned 'ceteris paribus' senses of preference, where one compares only the normal worlds of the relevant kinds.[38] Intriguingly, a source of similar ideas on entanglement is the semantics of expressions like "want" and "desire" in natural language, cf. [Stalnaker, 1984], [Heim, 1992], [Dandelet, 2014].

The preceding notions are similar to our earlier one with an intersection modality, but not quite. They only compare the two most plausible parts for each proposition.

We give no deeper analysis of all these entangled notions here, but as one small appetizer, we note that we are still within the bounds of this paper.

FACT 7 *$Pref^*$ is definable in a modal doxastic preference language.*

8.4 Preference change and belief revision

As we have observed already, our treatment of the statics and dynamics of belief and preference shows many similarities. It is an interesting test, then, if the earlier dynamic logic methods for pure cases transfer to belief-entangled notions of preference.

Intuitively, entangled preferences can change because of two kinds of trigger: evaluative acts like suggestions or commands, and informative acts changing our beliefs. As an illustration, we quote a result from [Liu, 2008]:

Theorem 25 *The dynamic logic of the above intersective preference H is axiomatizable, with the following essential recursion axioms:*

1. $\langle \sharp A\rangle\langle H\rangle\varphi \leftrightarrow (A \wedge \langle H\rangle(A \wedge \langle \sharp A\rangle\varphi)) \vee (\neg A \wedge \langle H\rangle\langle \sharp A\rangle\varphi)$.
2. $\langle \Uparrow A\rangle\langle H\rangle\varphi \leftrightarrow (A \wedge \langle H\rangle(A \wedge \langle \Uparrow A\rangle\varphi)) \vee (\neg A \wedge \langle H\rangle(\neg A \wedge \langle \Uparrow A\rangle\varphi)) \vee (\neg A \wedge \langle bett\rangle(A \wedge \langle \Uparrow A\rangle\varphi))$.

[37]Here, $B^{<}$ is an earlier-mentioned modality of *strong belief* that we do not define.

[38]This makes sense, for instance, in the field of epistemic game theory, where 'rationality' means comparing moves by their most plausible consequences according to the player's beliefs and then choosing the best.

3. $\langle A!\rangle\langle H\rangle\varphi \leftrightarrow A \wedge \langle H\rangle\langle A!\rangle\varphi$.

Having intersection modalities for static attitudes may not be all that is needed, though. Importantly, there may also be *entangled triggering events* that do not easily reduce to purely informational or purely evaluative actions, or sequential compositions thereof. Such entangled events, too, can be treated in our style, but we omit details here. [39]

Trade-offs between preference change and information change. Finally, as often in logic, distinctions can get blurred through redefinition. For instance, sometimes, the same scenario may be modeled either in terms of preference change, or as information change. Two concrete examples of such redescription are "Buying a House" in [de Jongh and Liu, 2009] and "Visit by the Queen" in [Lang and van der Torre, 2008]. Important though it is, we leave the study of precise connections between different representations of dynamic entangled scenarios to another occasion.

9 DEONTIC REASONING, CHANGING NORMS AND OBLIGATIONS

Our analysis of information and preference can itself be viewed as a study of normative discourse and reasoning. However, in this section, we turn to explicit deontic scenarios, and take a look at some major issues concerning obligations and norms from the standpoint of dynamic systems for preference change.[40]

9.1 Triggers for deontic actions and events

Perhaps the most immediate concrete task at hand as a testing ground for our treatment is charting the large variety of deontic notions in daily life. There is still an ongoing debate about identifying what are the major deontic notions and their meanings, witness the recent revival of interest in treating permission as a universal modality on its own in [Anglberger *et al.*, 2015], going back to early proposals in [Benthem, 1979]. Likewise, there is a large variety of dynamic deontic actions in daily life that affect normative attitudes and betterness orderings. Frequent normative triggers go far beyond the suggestions and commands that we chose as our examples. For instance, basic deontic acts also include the granting of permissions, [41] or the making of promises and threats – as should be clear from many chapters in this Handbook.

[39] For an analogy, see the question scenarios involving conversational triggers that induce parallel information and issue change in [Benthem and Minica, 2009].

[40] Our treatment largely follows [Benthem *et al.*, 2010] and [Benthem *et al.*, 2014].

[41] For some state of the art work on modeling permissions in deontic logic, see [Anglberger *et al.*, 2015]

We will not undertake an empirical survey of basic vocabulary here, since this is more of a task for moral philosophers or linguistic experts on normative discourse (see also the beginning of Section 10). Nevertheless, the examples in this chapter should have convinced the reader that a dynamic action perspective on deontic issues is natural, and that much can be done with the tools presented here.

Instead of engaging in further detailed studies of deontic discourse and reasoning, we merely consider a number of general topics and trends that have roots in the deontic literature.

9.2 Unary and dyadic obligation on ordering models

Our static logics heavily relied on binary ordering relations. In fact, deontic logic may have been the first area of philosophical logic to adopt this approach, building on observations from ethics that the deontic notions of obligation, permission and prohibition can be naturally made sense of in terms of an *ideality ordering* $\preceq$ on possible worlds. Here is a quote from [Moore, 1903], found in [Fraassen, 1973], p.6:

> " [...] to assert that a certain line of conduct is [...] absolutely right or obligatory, is obviously to assert that more good or less evil will exist in the world, if it is adopted, than if anything else be done instead."

In this line, the pioneering study [Hansson, 1969] interpreted dyadic obligations of the type 'it is obligatory that φ under condition ψ' on semantic models like ours, using a notion of maximality as in our study of belief:

$$\mathcal{M}, s \models O^{\psi}\varphi \Longleftrightarrow Max(||\psi||_{\mathcal{M}}) \subseteq ||\varphi||_{\mathcal{M}}$$

Depending on the properties of the relation $\preceq$, different deontic logics are obtained here: [Hansson, 1969] starts with a $\preceq$ which is only reflexive, moving then to total pre-orders. This is of course the same idea that has also emerged in conditional logic, belief revision, and the linguistic semantics of generic expressions.[42] Variations of this modeling have given rise to various preference-based semantics of deontic logic: see [van der Torre, 1997] for an early useful overview.

Recent developments show the continued vitality of this area. [Hansen, 2005] is a sophisticated study of conditional obligations in a setting of moral conflicts induced by promises or other deontic actions. Hansen's logic for conditional obligation uses van Fraassen-style reason-based deontic order

[42] A deontic criticism of this account has been that conditional obligation loses antecedent strengthening: [Tan and van der Torre, 1996]. This loss, however, makes sense in our view: non-monotonicity is inherent in the dynamics of information, where the set of most ideal worlds can change under update.

models while adding ideas reminiscent of 'premise semantics' in the area of conditional logic (as well as later strands in the semantics of non-monotonic logic), and it has a complete axiomatization using non-trivial techniques. Also noteworthy are [Parent, 2014] and [Parent, 2015] which settle several long-standing completeness questions for deontic logics using techniques from non-monotonic logic. This work also clarifies various options for defining 'maximality' and 'optimality' on conditional obligation, and shows how, in a deontic setting, some of the traditional technical fixes (such as the use of the 'limit assumption' in conditional semantics) can be circumvented, or at least, be analyzed in a more satisfactory manner.

In this light, our paper has taken up an old, but still active, strand in the semantics of deontic reasoning, and then added some recent themes concerning preference: criterion-based priority structure, dynamics of evaluative acts and events, and extended logical languages making these explicit. This seems a natural continuation of deontic logic, while also linking it up with developments in other fields.

9.3 Reasons and dynamics in classical deontic scenarios

The dynamic emphasis in this chapter on changes and their triggering events has thrown fresh light on the study of information and preference-based agency. Deontic logic proves to be no exception to this line of analysis, if we also bring in our treatment of reason-based preference (again we remind the reader of the pioneering [Fraassen, 1973]) – as we shall now demonstrate with a few examples.

The Gentle Murder scenario from [Forrester, 1984], p.194, is a classic of deontic logic that illustrates the basic problem of analyzing 'contrary-to-duty' obligations (CTDs).

Example 26 *"Let us suppose a legal system which forbids all kinds of murder, but which considers murdering violently to be a worse crime than murdering gently. [...] The system then captures its views about murder by means of a number of rules, including these two:*

1. *It is obligatory under the law that Smith not murder Jones.*
2. *It is obligatory that, if Smith murders Jones, Smith* [does so] *gently."*

The priority format of Section 6.1, even just with linear sequences, can represent this scenario in a natural way. Recall that a linear priority sequence $P_1, \ldots, P_n$ combines bipartitions $\{\mathcal{I}(p_i), -\mathcal{I}(p_i)\}$ of the domain of discourse S. Moving towards the right direction of the sequence, ever more atoms p_i are falsified. In a deontic reading, this means that, the more we move towards the right side of the sequence, the more violations hold of morally desirable properties.

Concretely, in the Gentle Murder scenario, the result is two classes of ideality: one class I_1 in which Smith does not murder Jones, i.e., $\mathsf{I}_1 := \neg m$; and another I_2 in which either Smith does not murder Jones or he murders him gently, i.e., $\mathsf{I}_2 := \neg m \vee (m \wedge g)$. The relevant priority sequence $\mathcal{B}$ has $\mathsf{I}_2 \prec \mathsf{I}_1$. Such a sequence orders the worlds via its induced relation $\preceq^{IM}_{\mathcal{B}}$ in three clusters. The most ideal states are those satisfying I_1, worse but not worst states satisfy $\mathsf{V}_1 := \neg\mathsf{I}_1$ but at the same time I_2, and, finally, the worst states satisfy $\mathsf{V}_2 := \neg\mathsf{I}_2$.

With this representation, we can take the scenario one step further.

Example 27 *Consider the priority sequence for Gentle Murder from the preceding Example: $\mathcal{B} = (\mathsf{I}_1, \mathsf{I}_2)$. We can naturally* restrict *$\mathcal{B}$ to an occurrence of the first violation by intersecting all formulas in the sequence with V_1. Then the first proposition becomes a contradiction, distinguishing no worlds. The best among the still available worlds are those with $Max^+(\mathcal{B}^{\mathsf{V}_1}) = \mathsf{I}_2 \wedge \mathsf{V}_1$. A next interesting restriction is $\mathcal{B}^{\mathsf{V}_2}$, which describes what the original priority sequence prescribes under the assumption that also the CTD obligation "kill gently" has been violated. In this case we end up in a set of states that are all equally bad.*

This brief sketch may suffice to show our approach provides a simple perspective on the deontic robustness of norms and laws viewed as *CTD* structures: they can still function when transgressions have taken place.[43]

Other major puzzles in the deontic literature, such as the Chisholm Paradox, are given similar reason-based representations in [Benthem *et al.*, 2014].

9.4 Typology of change at two levels

We have shown how two-level structure of preference provides a natural medium for modeling deontic notions. Likewise, it yields a rich account of deontic changes. In Section 7, we developed a theory of both informational and evaluative changes, operating either directly on possible world order, or on the priority structure underlying such orders. This two-pronged approach also makes sense here.

As an illustration, we add a temporal twist to the above classical deontic scenario, by 'dynamifying' Gentle Murder.

[43] Representing *CTD* structures in terms of chains of properties already occurs informally in [Fraassen, 1973]. A formal account is in [Governatori and Rotolo, 2005], with a Gentzen proof calculus manipulating formulae of the type $\varphi_1 @ \ldots @ \varphi_n$ with @ a connective representing a 'sub-ideality' relation. It is an interesting open problem if such a proof-theoretic approach can be related to the more semantically oriented modal logics of this chapter. Incidentally, the same sort of interface questions arise for more recent proof-theoretic approaches to deontic logic, such as the substructural logics for analyzing commands and permissions in [Anglberger *et al.*, 2014].

Example 28 *We start with a priority sequence $\mathcal{B} = (\neg m)$. This current deontic state of affairs generates a total pre-order where all $\neg m$ states are above all m states: "It is obligatory under the law that Smith not murder Jones". Now, we refine this order so as to introduce the sub-ideal obligation to kill gently: "it is obligatory that, if Smith murders Jones, Smith murders Jones gently". In other words, we want to model the process of refining legal codes, by introducing a contrary-to-duty obligation.*

Intuitively, this change can happen in one of two ways:

1. *We refine the given betterness ordering 'on the go' by requesting a further bipartition of the violation states, putting the $m \wedge g$-states above the $m \wedge \neg g$-states. This can be seen as the successful execution of a command of the sort "if you murder, then murder gently".*

2. *We introduce a new law 'from scratch', where $m \to g$ is now explicitly formulated as a class of possibly sub-ideal states. This can be seen as the enactment of a new priority sequence $(\neg m, m \to g)$.*[44]

The example illustrates how a CTD sequence can be dynamically created either by uttering a sequence of commands stating what ought to be the case in a sub-ideal situation, or by enacting a new priority sequence.

But in this setting, Theorem 4 from Section 7 applies: in terms of betterness among worlds, the two instructions amount to the same thing! In other words, in this scenario, the same deontic change can be obtained both by refining the order dictated by a given law, and by enacting a new law.

Of course, this is just a start, and not everything is smooth application. Our discussion of two-level dynamics in Section 7.2 and its possible failures of tracking also suggests new issues. For instance, some well-known changes in laws, such as *abrogation* (a counterpart to the earlier operation of 'graph deletion') have no obvious counterpart at the pure worlds level.

9.5 Norm change

The preceding discussion leads up to a more general theme of global dynamics. The problem of *norm change* has recently gained attention from researchers in deontic logic, legal theory, as well as multi-agent systems.

Approaches to norm change fall into two groups. In syntactic approaches—inspired by legal practice—norm change is an operation performed directly on the explicit provisions in the code of the normative system [Governatori and Rotolo, 2008a], [Governatori and Rotolo, 2008b], [Boella *et al.*, 2009]. In semantic approaches, however, norm change tends to follow deontic preference order (cf. [Aucher *et al.*, 2009b]). Our initial betterness dynamics

[44] We have encountered this pattern before, since $m \to g$ is equivalent to $\neg m \vee (m \wedge g)$.

on models belonged to the latter group, but our priority methods tie it to the former more syntactic level of representation. [45]

More drastic changes of norms and moral codes can be modeled, too, in our framework, using the calculus of priority graphs that we have sketched in Section 6. For an extended discussion of norm change and even legal code change, we refer again to [Benthem *et al.*, 2014].

9.6 Entangled changes

Finally, as observed already in Section 8 on entanglement (cf. [Lang *et al.*, 2003] for a deontic discussion), the dynamic logic connection allows for a unified treatment of two kinds of change that mix harmoniously in deontic scenarios: information change given a fixed normative order, and evaluation change modifying such an order.

Natural deontic scenarios can have deeply intertwined combinations of obligation, knowledge and belief. This point has been acknowledged in the recent literature, and led to combinations of deontic and epistemic modalities, [Aucher *et al.*, 2011], [Balbiani and Seban, 2011]. In such a setting, simple operator combinations already express intriguing notions, witness distinctions such as that between KO ('knowing one's duty') versus OK ('having a duty to know'). We add a few more illustrations.

The first example comes from [Liu, 2011a]. Let us consider the conditional obligation $O^{\psi}\varphi$ again. We can understand the condition ψ as a fact, then $O^{\psi}\varphi$ would express an obligation based on somewhat objective condition. However, fullfilling obligations unavoidably involves agents, hence their epistemic attitude immediately become relevant. With this spirit, we can take the condition at least in the following two sense: (a) an agent *knows* that ψ is true or (b) an agent *believes* that ψ is true. In the former case, we would get much weaker obligation, which in constrast with the stronger obligation obtained in the latter case.

Some sophisticated moral scenarios in [Pacuit *et al.*, 2006] go even further than simple combination, and point at the further conceptual subtleties arising in a dynamic setting congenial to our main theme in this chapter. These include the distinction between learning new facts that trigger duties, such as accidentally finding out that my neighbor is in distress, or having a duty to know, as happens with the intensive care department of a hospital that is supposed to know the condition of their patients. These issues are interesting and worth pursuing. As far as we know, there has been no sustained systematic analysis yet following up on this work.

Many further deontic themes can be analyzed along the above lines. We refer to [Benthem *et al.*, 2010], [Benthem *et al.*, 2014] for a detailed treat-

[45] The bridge here is our earlier analysis: obligations defined via ideality and maximality are special kinds of classifications of an Andersonian-Kangerian type.

ment of the Chisholm Paradox, and concrete ways in which priority graph calculus models norm change.

Summary. Taken together, the themes presented in this section show how the logical perspective of this chapter connects with deontic issues, and can throw new light on them. Admittedly, for our style of analysis to work, we do need a few ingredients that are not part of traditional formalizations in deontic logic: in particular, dynamic events, and reasons underlying world ordering. But we believe that such ingredients are there in the very examples used in the field, provided that we 'mine' their texts for additional dynamic and criterion-based linguistic cues, or just re-analyze the relevant deontic scenarios in these richer terms. The above illustrations may at least have suggested that, and how, this might be done.

10 FURTHER DIRECTIONS

Our main presentation has come to an end. Even so, many relevant roads lead from here. Collecting some points from earlier sections, here are a few active directions where deontic logic meets, or could meet, with current trends in dynamic logics of agency.

10.1 Language, speech acts, and agency

Events that drive information or preference change are often *speech acts* of telling, asking, and so on. Natural language has a sophisticated repertoire of speech acts with a deontic flavour (commanding, promising, allowing, and so on) that invite further logical study, taking earlier studies in meta-ethics and Speech Act Theory (cf. [Searle and Veken, 1985]) to the next level. In particular, such studies will also need a more fine-grained account of the *multi-agency* in dynamic triggers, that has been ignored in this chapter. For instance, things are said by someone to someone, and their uptake depends on relations of authority or trust. Likewise, promises, commands, or permissions are given by someone to someone, and their normative effect depends in subtle ways on who does, and is, what. In particular, [Yamada, 2010] is a pioneering study of this fine-structure of normative action using dynamic-epistemic logic.

10.2 Multi-agency and groups

A conspicuous turn in studies of information dynamics has been a strong emphasis on social scenarios with multi-agent interaction. After all, language use is about communication between different agents, a major paradigm for logic is argumentation between different parties, social behaviour is kept in place by mutual expectations, and so on. In the logics for knowledge, belief, and preference of this chapter, this multi-agent turn can be represented by iteration of single-agent modalities, as in a's knowing that b does, or does not, knows some fact, [Benthem, 2011]. The same is true for games (cf. [Benthem, 2014]), a topic we will address later.

However, eventually, in a social setting, *groups* must also be taken seriously as new collective actors in their own right. Then we need logics that can deal with notions such as 'common knowledge' or 'distributed knowledge', and their counterparts for beliefs (cf. [Fagin *et al.*, 1995], [Meyer and van der Hoek, 1995], [Baltag *et al.*, 2015]), or even with more truly collective group-level preferences in Social Choice Theory (cf. [Endriss, 2011]).

All these logics of social behavior have dynamic-epistemic extensions in the style of this chapter, and state-of-the-art samples may be found in [Seligman *et al.*, 2013], [Baltag *et al.*, 2013], [Hansen and Hendricks, 2014].

The social turn is highly relevant to deontic logic. From the start, deontic notions and morality seems all about *others*: my duties are usually toward other people, my norms come from outside sources: my boss or a lawgiver.[46] In principle, the methods of this paper can deal with social multi-agent structure in deontic settings, though much remains to be understood. For instance, it is easy to interpret informational iterations such as K_aK_bp, in volving different agents – but what, for instance, is the meaning of an iterated obligation O_aO_bp? And beyond this, what would be a group-based 'common obligation': is this more like a propositional common belief, or like a demand for joint action of the group? Other relevant issues in this setting are the entanglement of informational and evaluative acts for groups: cf. [Hartog, 1985], [Kooi and Tamminga, 2006], [Konkka, 2000], and [Holliday, 2009] on morality as held together by social expectations such as trust. An account of deontically relevant actions for groups will also have to include new operations reminiscent of social choice, such as *belief merge* and *preference merge*, where the priority structures of Section 6 may find a new use as a model for social institutions: cf. [Grossi, 2007].

10.3 Games and dependent behavior

Multi-agency involves not just social knowledge, beliefs, and preferences, but also by individual and collective action. All these notions come together concretely in the area of *games*, and hence, not surprisingly, logics of agency have close connections with game theory ([Shoham and Leyton-Brown, 2008], [Benthem, 2014]), being the general mathematical study of strategic behavior and its equilibria.

In the normative realm, actions are as crucial as states of the world – even though actions have been largely ignored in this chapter, for reasons of

[46]This social aspect has been clearly acknowledged by computer scientists working on multi-agent systems: cf. [Meyer, 1988], [Wooldridge, 2000], and [Rao and Georgeff, 1991].

space. In particular, dependent action is crucial in deontic practice (think of sanctions or rewards), and games are a congenial paradigm. Indeed, many topics in this chapter suggest game-theoretic extensions. In particular, we saw how belief-entangled set lifting is crucial to rational choices made by agents, and this entanglement is typical for games. Thus, multi-agent versions of our logics have turned out to be a natural tool in the analysis of game solution procedures (cf. [Roy, 2008], [Dégremont, 2010], [Benthem, 2014]). But preference change also makes sense in games, once we see their preference structure not as a static given, but as something that can evolve dynamically during play. For some first excursions in this direction, see [Liu, 2011a] on the topic of rationalizing preferences in the curse of, or after, playing a game.

Another intriguing line worth mentioning are recent uses of deontic logic as a sort of high-zoom level language for our ordinary discourse about 'optimal action', where the precise details of game-theoretic solution procedures such as Backward Induction have been suppressed: cf. [Benthem *et al.*, 2006], [Benthem, 2014], and [Roy *et al.*, 2014]. This may well become a major new interpretation for deontic formalisms.

But there is also a converse direction in this contact. Ideas from game theory have started entering deontic logic. One interesting example is the use of standard game solution methods as deliberation procedures for moral judgments in [Loohuis, 2009] and [Tamminga, 2013]. One might even argue that dependent social behavior is the very source of morality, and in that sense, games would be a mandatory next stage after the single-episode driven dynamic logics of this paper.

10.4 Temporal perspective

Games are one longer-term activity, but deontic agency involves many different processes, some even infinite. The general logical setting here are temporal logics (cf. [Fagin *et al.*, 1995], [Parikh and Ramanujam, 2003]) where new phenomena come to the fore. Deontics and morality is not just about single episodes, but about action and interaction over time. Early work in deontic logic already used temporal logics: cf. the pioneering dissertation [Eck, 1981]) where events happen in infinite histories, and obligations come and go. Likewise, in the multi-agent community, logics have been proposed for preferences between complete histories, and planning behavior leading to most desired histories (cf. [Meyden, 1996], [Sergot, 2004]). Such temporal logics mesh well with dynamic-epistemic logics (cf. [Benthem *et al.*, 2009a]), with an interesting role for *protocols* as a new object of study, i.e., available procedures for reaching goals. Plans and protocols have a clear normative dimension as well, and thus one would wish to incorporate them into the preference dynamics of this paper.

Our logics described single, or just a few, update or reasoning steps, and the same is true for most scenarios in the deontic literature. However, the broader horizon of single steps is the temporal process of inquiry in the informational case, and the long-term functioning of society in the normative case. Studying both aspects together, local and global, seems essential.

10.5 Fine-structure of information

Most dynamic logics for agency, whether about information dynamics or evaluation dynamics, are semantic in nature. The states changed by the process are semantic models. However, in philosophical logic, there has been a continuing debate about the right representation of the *information* used by agents. Semantic information as used in this chapter, though common to many areas, including decision theory and game theory, is coarse-grained, identifying logically equivalent propositions, making agents 'omniscient' at least to that degree – thereby suppressing the very activity of logical analysis as an information-producing process.

Zooming in on the latter dynamics, agents engage in many activities, such as inference, memory retrieval, introspection, or other forms of 'awareness management' that require a more fine-grained notion of information, closer to syntax. Several dynamic logics of this kind have been proposed in recent years, using ideas from proof theory rather than model theory: cf. [Jago, 2006], [Velazquez-Quesada, 2009], and the survey chapter [Benthem and Martínez, 2008] on different notions of information in modern logic. [47]

But new levels keep appearing. One compromise are the 'evidence models' of [van Benthem & Pacuit, 2011] that generalize the modal logics of this chapter to a 'neighborhood semantics' recording the evidence generating the plausibility ordering on which our modeling of belief was based. While this intermediate level still identifies equivalent propositions in the sense of its weaker base logic, it turns out to support a much richer account of events triggering evidence change: closer, in some ways, to the dynamics of our earlier priority graphs.

Finally, these various levels of representing information are not at odds with each other. Another recent topic is that of 'tracking', cf. [Benthem, 2016b] and [Baltag *et al.*, 2016], mentioned several times already, where one studies systematically under which conditions updates at a coarser level can faithfully track updates performed at some finer level.

The same issues of grain level for information make sense in the deontic realm. For instance, our priority graphs with reasons for preferences were syntactic objects than get manipulated by insertions, deletions, permutations, and the like. Significantly, 'reason' is a proof- or argumentation-based

[47] The latter distinguishes even further natural varieties of information that can be found in logic today, such as 'correlation-based' and 'procedural' information.

term. And indeed, deontic logic has more fine-grained proof-theoretic aspects that would be swept under the rug in a purely semantic approach. As just one illustration, consider the following obvious counterpart to the above-mentioned problem of omniscience. My moral obligations to you cannot reasonably be based on my foreseeing every consequence of my duties or commitments. I owe you careful deliberation, not omniscience. [48] For this and other reasons, there is room for more fine-grained dynamic representations of information and evaluation, closer to deontic syntax – where model theory and proof theory may find interesting ways to meet, for instance, [Tosatto *et al.*, 2012], and [Dong and Gratzl, 2016]

10.6 Digression: numerical strength

While the main theme of this chapter is qualitative approaches, it should be mentioned that there are also numerical approaches to preferences, employing utilities (cf. [Rescher, 1966], [Trapp, 1985]) or more abstract 'grades' for worlds (cf. [Spohn, 1988]). Dynamic ideas work in this setting, too, witness the modal logic with graded modalities indicating the strength of preference in [Aucher, 2003], which also defines product update for numerical plausibility models. A stream-lined version in [Liu, 2004] uses propositional constants q_a^m saying that agent a assigns the current world a value of at most m. Our earlier ordering models, both for plausibility and for preference, now get numerical graded versions, with more finely-grained statements of strength of belief and of preference. Dynamic updates can now be defined where we assign values to actions or events, using numerical stipulations in terms of 'product update' from the cited references.[49] More complex numerical evaluation uses *utility* as a fine-structure of preference, and its dynamics can also be dealt with in this style: cf. [Liu, 2004], [Liu, 2009b].

While the technical details of these approaches are not relevant here, systems like this do address two issues that seem of great deontic relevance. One is the possibility of comparing not just worlds qua preference, but also *actions*, making sense of the principled distinction in ethics between outcome-oriented and deontological views of obligations and commitments. The other major benefit of a quantitative approach is that we can now study the logic of *how much good* an action does, and accordingly, measure the extent to which we can improve current situations by our actions.

[48]Likewise, citizens are supposed to know the law, but it would be both unrealistic and unfair to require them to be as well-versed as professional lawyers.

[49]The resulting dynamic logic of numerical evaluation can be axiomatized in the same recursive style as the qualitative systems we discussed in this chapter.

10.7 Probability

Another natural quantitative addition to our analysis would be *probability.* Probabilities measure strengths of beliefs, thereby providing fine-structure to the plausibility orderings that we have worked with. But they can also indicate information that we have about a current process, or a reliability we assign to our observation of a current event.[50] Finally, the numerical factors in probability theory also allow us to mix and weigh various factors in the entangled versions of preference and deontic notions discussed in Section 8. A striking entangled notion is *expected value* in probability theory, whose definition mixes beliefs and evaluation. A unified treatment of logical and probabilistic perspectives in the deontic realm seems a clear desideratum.

11 APPENDIX: RELEVANT STRANDS IN THE LITERATURE

The themes of this paper have a long history. For instance, we have pointed at the important connections with belief revision theory and non-monotonic logics throughout. Moreover, while we have followed the dynamic-epistemic approach, there are other proposals in the literature for combining and 'dynamifying' preferences, beliefs, and obligations. In addition to the literature cited already, here are some other relevant lines of work that could not fit into the main line of our presentation.

Computation and agency. [Meyer, 1988] is a pioneering study of deontics from a dynamic viewpoint, reducing deontic logics to suitable dynamic logics. In the same tradition, [Meyden, 1996] takes the deontic logic/dynamic logic interface a step further, studying 'free choice permission' with a new dynamic logic where preferences can hold between actions. Completeness theorems for this enriched semantics then result for several systems. [Pucella and Weissmann, 2004] provide a dynamified logic of permission that builds action policies for agents by adding or deleting transitions. [Demri, 2005] reduces an extension of van der Meyden's logic to PDL, yielding an EXPTIME decision procedure, and showing how PDL can deal with agents' policies. Preference semantics has also been widely used in AI tasks: e.g., [Wellman and Doyle, 1991] gives a preference-based semantics for goals in decision theory. This provides criteria for verifying the design of goal-based planning strategies, and a new framework for knowledge-level analysis of planning systems. [Horty, 1993] studies commonsense normative reasoning, arguing that techniques of non-monotonic logic provide a better framework than the usual modal treatments. The paper has applications to conflicting obligations and conditional obligations. [Lang *et al.*, 2003] propose a logic of desires whose semantics contains two ordering relations of preference and normality, and then interpret "in context A, I desire B" as 'the best among

[50] See [Benthem *et al.*, 2009b] for a rich dynamic epistemic logic of probability.

the most normal $A \wedge B$ worlds are preferred to the most normal $A \wedge \neg B$ worlds', providing a new entanglement of preference and normality.

Semantics of natural language. In a line going back to [Spohn, 1988], [Veltman, 1996] presents an update semantics for default rules, locating their meaning in the way in which they modify expectation patterns. This is part of a general program of 'update semantics' for conditionals and other key expressions in natural language. [van der Torre and Tan, 1999] use ideas from update semantics to formalize deontic reasoning about obligations. In their view, the meaning of a normative sentence resides in the changes it brings about in the 'ideality relations' of agents to whom a norm applies. [Zarnic, 2003] uses a simple dynamic update logic to formalize natural language imperatives of the form $FIAT\ \varphi$, which can be used in describing the search for solutions of planning problems. [Mastop, 2005] extends the update semantic analysis of imperatives to include third person and past tense imperatives, while also applying it to the notion of free choice permission. [Parent, 2003] outlines a preference-based account of communication, which brings the dynamics of changing obligations for language users to the fore. [Yamada, 2008] distinguishes the illocutionary acts of commanding from the perlocutionary acts that affect preferences of addressees, proposes a new dynamic logic which combines preference upgrade and deontic update, and discusses some deontic dilemmas in this setting.

Philosophical logic. The philosophical study of agency has many themes that are relevant to this paper, often inspired by topics in epistemology or by the philosophy of action. In a direction that is complementary to ours, with belief change as a starting point, [Hansson, 1995] identifies four types of changes in preference, namely revision, contraction, addition and subtraction, and shows that they satisfy plausible postulates for rational changes. The collection [Grune-Yanoff and Hansson, 2009] brings together the latest approaches on preference change from philosophy, economics and psychology. Following Hansson's work, [Alechina *et al.*, 2013] defines minimal preference change in the spirit of AGM framework and characterises minimal contraction by a set of postulates. A linear time algorithm is proposed for computing preference changes. In addition, going far beyond what we have discussed in this chapter, Hansson has written a series of seminal papers combining ideas from preference logic and deontic logic, see e.g. [Hansson, 1990b], [Hansson, 1990a] and [Hansson, 2001b].

Rational choice theory. Preference is at the heart of decision and rational choice. In recent work at the interface of preference logic, philosophy, and social science, themes from our chapter such as reason-based and belief-entangled preference have come to the fore, with further lines of their own. [Dietrich and List, 2013b] and [Dietrich and List, 2013a] point out that, though existing decision theory gives a good account of how agents make choices given their preferences, issues of where these essential preferences come from and how they can change are rarely studied. [51] The authors propose a model in which agents' preferences are based on 'motivationally salient properties' of alternatives, consistent sets of which can be compared using a 'weighing relation'. Two intuitive axioms are identified in this setting that precisely characterize the property-based preference relations. Starting from similar motivations, [Osherson and Weinstein, 2012a] studies reason-based preference in more complex doxastic settings, drawing on ideas from similarity-based semantics for conditional logic. Essentially, preference results here from agents' comparing two worlds, one having some property and the other lacking it, close to their actual world, and comparing these based on relevant aspects of utility. The framework supports extensive analysis in modal logic, including illuminating results on frame correspondence and axiomatization. [Osherson and Weinstein, 2012b] gives an extension to preference in the presence of quantifiers, while [Osherson and Weinstein, 2014] makes a link between these preference models and deontic logic.

12 CONCLUSION

We have shown how dynamic-epistemic logics can deal with information, knowledge, belief, but also with preference, intrinsic or based on criteria, as well as changes in all these dimensions as events happen and agents act. In doing so, we obtained a suggestive framework for the analysis of deontic notions that links them with many strands in the literature on agency. We also hope to have shown how pursuing this perspective may yield a fresh look at many existing normative scenarios, and may suggest new technical questions about deontic logic as traditionally conceived.

BIBLIOGRAPHY

[Alechina *et al.*, 2013] N. Alechina, F. Liu, and B. Logan. Minimal preference change. In D. Grossi, O. Roy, and H. Huang, editors, *Proceedings of the 4th International Workshop on Logic, Rationality and Interaction (LORI 2013)*, volume 8196 of *FoLLI-LNCS*, pages 15–26. Springer, 2013.

[Andréka *et al.*, 2002] H. Andréka, M. Ryan, and P-Y. Schobbens. Operators and laws for combining preferential relations. *Journal of Logic and Computation*, 12:12–53, 2002.

[Anglberger *et al.*, 2014] A. Anglberger, H. Dong, and O. Roy. Open reading without free choice. In F. Cariani, D. Grossi, J. Meheus, and X. Parent, editors, *Proceedings of 12th International Conference (DEON 2014)*, 2014, pp.19–32.

[Anglberger *et al.*, 2015] A. Anglberger, N. Gratzl, and O. Roy. Obligation, free choice and the logic of weakest permissions. *The Review of Symbolic Logic*, 8:807–827, 2015.

[Åqvist, 1987] L. Åqvist. *Introduction to Deontic Logic and the Theory of Normative Systems*. Naples: Bibliopolis, 1987.

[Åqvist, 1994] L. Åqvist. Deontic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 605–714. Dordrecht: Kluwer, 1994.

[51] These are of course precisely the two main topics of this paper: cf. also [Liu, 2011a].

[Aucher *et al.*, 2009a] G. Aucher, P. Balbiani, LF. del Cerro, and A. Herzig. Global and local graph modifiers. *Electronic Notes in Theoretical Computer Science*, 231:293–307, 2009.

[Aucher *et al.*, 2009b] G. Aucher, D. Grossi, A. Herzig, and E. Lorini. Dynamic context logic. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 15–26. Springer, 2009.

[Aucher *et al.*, 2011] G. Aucher, G. Boella, and L. van der Torre. A dynamic logic for privacy compliance. *Artif. Intell. Law*, 19(2-3):187–231, 2011.

[Aucher *et al.*, 2016] G. Aucher, J. van Benthem, and D. Grossi. Sabotage modal logic revisited. To appear in LORI special issue, 2016.

[Aucher, 2003] G. Aucher. A combined system for update logic and belief revision. Master's thesis, MoL-2003-03. ILLC, University of Amsterdam, 2003.

[Balbiani and Seban, 2011] P. Balbiani and P. Seban. Reasoning about permitted announcements. *Journal of Philosophical Logic*, 40(4):445–472, 2011.

[Baltag and Smets, 2008] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In M. Wooldridge G. Bonanno, W. van der Hoek, editor, *Logic and the Foundations of Game and Decision Theory*, volume 3 of *Texts in Logic and Games*. Amsterdam: Amsterdam University Press, 2008.

[Baltag *et al.*, 1998] A. Baltag, L.S. Moss, and S. Solecki. The logic of common knowledge, public announcements, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, pages 43–56, 1998.

[Baltag *et al.*, 2009] A. Baltag, S. Smets, and J. Zvesper. Keep 'hoping' for rationality: A solution to the backward induction paradox. *Synthese*, 169(2):301–333, 2009.

[Baltag *et al.*, 2011] A. Baltag, N. Gierasimczuk, and S. Smets. Belief revision as a truth tracking process. In K.R. Apt, editor, *Proceedings of Theoretical Aspects of Rationality and Knowledge (TARK 2011)*, 2011. ACM Digital Library, pp.187–190.

[Baltag *et al.*, 2013] A. Baltag, Z. Christoff, J.U. Hansen, and S. Smets. Logical models of informational cascades. In J. van Benthem and F. Liu, editors, *Logic Across the University: Foundations and Application*, pages 405–432. College Publications, London, 2013.

[Baltag *et al.*, 2015] A. Baltag, N. Bezhanishvili, A. Özgün, and S. Smets. The topological theory of belief. ILLC Research Report PP-2015-18, 2015.

[Baltag *et al.*, 2016] A. Baltag, J. van Benthem, and G.Ciná. Tracking tracking with categories. working paper, ILLC, University of Amsterdam, 2016.

[Benthem and Liu, 2007] J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17:157–182, 2007.

[Benthem and Martínez, 2008] J. van Benthem and M. Martínez. The stories of logic and information. In J. van Benthem and P. Adriaans, editors, *Handbook of Philosophy of Information*. Amsterdam: Elsevier, 2008.

[Benthem and Minica, 2009] J. van Benthem and S. Minica. Toward a dynamic logic of questions. In X. He, J. F. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 27–41. Springer, 2009.

[Benthem and Pacuit, 2011] J. van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99:61–92, 2011.

[Benthem and Smets, 2015] J. van Benthem and S. Smets. Dynamic logics of belief change. In H. van Ditmarsch, J.Y. Halpern, W. van der Hoek, and B. Kooi, editors, *Handbook of Logics for Knowledge and Belief*, pages 313–393. College Publications, 2015.

[Benthem *et al.*, 2006] J. van Benthem, S. van Otterloo, and O. Roy. Preference logic, conditionals and solution concepts in games. In H. Lagerlund, S. Lindström, and R. Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, pages 61–77. Uppsala Philosophical Studies 53, 2006.

[Benthem *et al.*, 2009a] J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, 2009.

[Benthem *et al.*, 2009b] J. van Benthem, J. Gerbrandy, and B. Kooi. Dynamic update with probabilities. *Studia Logica*, 93(1):67–96, 2009.

[Benthem *et al.*, 2009c] J. van Benthem, P. Girard, and O. Roy. Everything else being equal: A modal logic approach for ceteris paribus preferences. *Journal of Philosophical Logic*, 38(1):83–125, 2009.

[Benthem *et al.*, 2010] J. van Benthem, D. Grossi, and F. Liu. Deontics = betterness + priority. In G. Governatori and G. Sartor, editors, *Deontic Logic in Computer Science, 10th International Conference, DEON 2010*, volume 6181 of *LNAI*, pages 50–65. Springer, 2010.

[Benthem *et al.*, 2014] J. van Benthem, D. Grossi, and F. Liu. Priority structures in deontic logic. *Theoria*, 80(2):116–152, 2014.

[Benthem, 1979] J. van Benthem. Minimal deontic logics. *Bulletin of the Section of Logic*, 8:36–41, 1979.

[Benthem, 2006] J. van Benthem. Belief update as social choice. In P. Girard, O. Roy, and M. Marion, editors, *Dynamic Formal Epistemology*, pages 151–160. Springer, Dordrecht, 2006.

[Benthem, 2007] J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logic*, 17:129–156, 2007.

[Benthem, 2010] J. van Benthem. *Modal Logic for Open Minds*. Stanford: CSLI Publications, 2010.

[Benthem, 2011] J. van Benthem. *Logical Dynamics Information And Interaction*. Cambridge University Press, 2011.

[Benthem, 2014] J. van Benthem. *Logic in Games*. The MIT Press, 2014.

[Benthem, 2016a] J. van Benthem. Talking about knowledge. In L. Moss C. Başkent and R. Ramanujam, editors, *Rohit Parikh on Logic, Language and Society*. Springer, 2016.

[Benthem, 2016b] J. van Benthem. Tracking information. ILLC Research Report PP-2016-02, 2016.

[Blackburn *et al.*, 2001] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge: Cambridge University Press, 2001.

[Blackburn *et al.*, 2007] P. Blackburn, J. van Benthem, and F. Wolter. *Handbook of Modal Logic*. Elsevier, 2007.

[Boella *et al.*, 2009] G. Boella, G. Pigozzi, and L. van der Torre. Normative framework for normative system change. In P.Decker, J.Sichman, C.Sierra, and C.Castelfranchi, editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 169–176, 2009.

[Boutilier, 1992] C. Boutilier. *Conditional Logics for Default Reasoning and Belief Revision*. PhD thesis, University of Toronto, 1992.

[Boutilier, 1994] C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68:87–154, 1994.

[Brandom, 1994] R. Brandom. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1994.

[Burgess, 1984] J. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 89–133. Dordrecht: D. Reidel, 1984.

[Coste-Marquis *et al.*, 2004] S. Coste-Marquis, J. Lang, P. Liberatore, and P. Marquis. Expressive power and succinctness of propositional languages for preference representation. In D. Dubois, C. Welty, and M-A. Williams, editors, *Proceedings of the 9th International Conference on Principles of Knowledge Representation and Reasoning (KR 2004)*. Menlo Park, CA: AAAI Press, 2004. pp.203–212.

[Dong and Gratzl, 2016] H. Dong and N. Gratzl. Open reading for free choice permission: A perspective in substructural logics. Manuscript, 2016.

[Dandelet, 2014] S. Dandelet. Partial desires, blinkered beliefs. Manuscript, January 22, 2014.

[de Jongh and Liu, 2009] D. de Jongh and F. Liu. Preference, priorities and belief. In T.Grune-Yanoff and S.O. Hansson, editors, *Preference Change: Approaches from Philosophy, Economics and Psychology*, Theory and Decision Library, pages 85–108. Springer, 2009.

[Dégremont, 2010] C. Dégremont. *The Temporal Mind. Observations on the Logic of Belief Change in Interactive Systems.* PhD thesis, ILLC, University of Amsterdam, 2010.

[Demri, 2005] S. Demri. A reduction from DLP to PDL. *Journal of Logic and Computation*, 15:767–785, 2005.

[Dietrich and List, 2013a] F. Dietrich and C. List. A reason-based theory of rational choice. *Nous*, 47(1):104–134, 2013.

[Dietrich and List, 2013b] F. Dietrich and C. List. Where do preferences come from? *International Journal of Game Theory*, 42(3):613–637, 2013.

[Eck, 1981] J. van Eck. *A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications.* PhD thesis, University of Groningen, 1981.

[Endriss, 2011] U. Endriss. Applications of logic in social choice theory. In J. Leite, P. Torroni, Th. Agotnes, G. Boella, and L.van der Torre, editors, *Proceedings of the 12th International Workshop on Computational Logic in Multiagent Systems (CLIMA-2011)*, volume 6814 of *LNAI*, pages 88–91. Springer-Verlag, July 2011. Extended abstract corresponding to an invited talk.

[Fagin *et al.*, 1995] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge.* Cambridge, MA: The MIT Press, 1995.

[Forrester, 1984] J. Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81:193–197, 1984.

[Fraassen, 1972] B. van Fraassen. The logic of conditional obligation. *Journal of Philosophical Logic*, 1:417–438, 1972.

[Fraassen, 1973] B. van Fraassen. Values and the heart's command. *The Journal of Philosophy*, 70(1):5–19, 1973.

[Friedman and Halpern, 1995] N. Friedman and J. Halpern. Plausibility measures: A user's guide. In P. Besnard and S. Hanks, editors, *Proc. Eleventh Conf. on Uncertainty in Artificial Intelligence (UAI 95)*, Morgan Kaufmann, San Francisco, USA, 1995. pages 175–184.

[Friedman and Halpern, 1997] N. Friedman and Joseph Y. Halpern. Modeling belief in dynamic systems, part I: Foundations. *Artificial Intelligence*, 95(2):257–316, 1997.

[Friedman and Halpern, 1999] N. Friedman and Joseph Y. Halpern. Modeling belief in dynamic systems. part II: Revision and update. *Journal of Artificial Intelligence Research*, 10:117–167, 1999.

[Gabbay *et al.*, 2013] D. Gabbay, J.Horty, X. Parent, R. van der Meyden, and L. van der Torre. *Handbook on Deontic Logic and Normative Systems*, 2013. College Publications: London.

[Gerbrandy, 1999] J. Gerbrandy. *Bisimulation on Planet Kripke.* PhD thesis, ILLC, University of Amsterdam, 1999.

[Girard *et al.*, 2012] P. Girard, J. Seligman, and F. Liu. General dynamic dynamic logic. In Thomas Bolander, Torben Braüner, Silvio Ghilardi, and Lawrence S. Moss, editors, *Advances in Modal Logic*, pages 239–260. College Publications, 2012.

[Girard, 2008] P. Girard. *Modal Logics for Belief and Preference Change.* PhD thesis, Stanford University, 2008.

[Goble, 2000] L. Goble. Multiplex semantics for deontic logic. *Nordic Journal of Philosophical Logic*, 5(2):113–134, 2000.

[Governatori and Rotolo, 2005] G. Governatori and A. Rotolo. Logic of violations: A gentzen system for reasoning with contrary-to-duty obligations. *Australasian Journal of Logic*, 3:193–215, 2005.

[Governatori and Rotolo, 2008a] G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. Part 1: Revision and defeasible theories. In R. van der Meyden and L. van der Torre, editors, *Proceedings of the 9th International Conference on Deontic Logic in Computer Science (DEON 2008)*, volume 5076 of *LNAI*, pages 3–18. Springer, 2008.

[Governatori and Rotolo, 2008b] G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. Part 2: Temporalised defeasible logic. In G. Boella, G. Pigozzi, M. P. Singh, and H. Verhagen, editors, *Proceedings of the 3rd International Workshop on Normative Multiagent System (NorMAS 2008), Luxembourg, Luxembourg, July 14-15, 2008*, pages 112–127, 2008.

[Grandi and Endriss, 2009] U. Grandi and U. Endriss. First-order logic formalisation of Arrow's theorem. In *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI-2009)*, volume 5834 of *LNAI*, pages 133–146. Springer-Verlag, October 2009. Also presented at DGL-2009.

[Grossi, 2007] D. Grossi. *Desigining Invisible Hancuffs. Formal Investigations in Institutions and Organizations for Multi-Agent Systems.* PhD thesis, Utrecht University, 2007. SIKS Dissertation Series 2007-16.

[Grove, 1988] A. Grove. Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[Grune-Yanoff and Hansson, 2009] T. Grune-Yanoff and S.O. Hansson, editors. *Preference Change: Approaches from Philosophy, Economics and Psychology.* Theory and Decision Library. Springer, 2009.

[Guo and Sliegman, 2011] M. Guo and J. Sliegman. Making choices in social situations. In *Logic and Interactive Rationalityok*, pages 176–202. ILLC, University of Amsterdam, 2011.

[Halpern, 1997] J.Y. Halpern. Defining relative likelihood in partially-ordered preferential structure. *Journal of Artificial Intelligence Research*, 7:1–24, 1997.

[Hansen and Hendricks, 2014] P. G. Hansen and V. F. Hendricks. *Infostorms: How to Take Information Punches and Save Democracy.* Copernicus Books / Springer, 2014.

[Hansen, 2005] J. Hansen. Conflicting imperatives and dyadic deontic logic. *Journal of Applied Logic*, 2:484–511, 2005.

[Hansson, 1969] B. Hansson. An analysis of some deontic logics. *Nous*, 3:373–398, 1969.

[Hansson, 1990a] S.O. Hansson. Defining 'good' and 'bad' in terms of 'better'. *Notre Dame of Journal of Formal Logic*, 31:136–149, 1990.

[Hansson, 1990b] S.O. Hansson. Preference-based deontic logic. *Journal of Philosophical Logic*, 19:75–93, 1990.

[Hansson, 1995] S.O. Hansson. Changes in preference. *Theory and Decision*, 38:1–28, 1995.

[Hansson, 2001a] S.O. Hansson. Preference logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 319–393. Dordrecht: Kluwer, 2001.

[Hansson, 2001b] S.O. Hansson. *The Structure of Values and Norms.* Cambridge: Cambridge University Press, 2001.

[Hartog, 1985] G. den Hartog. *Wederkerige Verwachtingen [Mutual Expectations].* PhD thesis, University of Amsterdam, 1985.

[Heim, 1992] I. Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221, 1992.

[Holliday, 2009] W.H. Holliday. Dynamic testimonial logic. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality, and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 161–179. Springer, 2009.

[Holliday, 2012] W.H. Holliday. *Knowing what follows: epistemic closure and epistemic logic.* PhD thesis, Stanford University, 2012.

[Holliday, 2014] W.H. Holliday. Epistemic closure and epistemic logic I: Relevant alternatives and subjunctivism. *Journal of Philosophical Logic*, 2014.

[Horty, 1993] J. Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9:69–91, 1993.

[Houser and Kurzban, 2002] D. Houser and R. Kurzban. Revealed preference, belief, and game theory. *Economics and Philosophy*, 16:99–115, 2002.

[Icard III *et al.*, 2010] T. Icard III, E. Pacuit, and Y. Shoham. Joint revision of beliefs and intentions. In F. Lin, U. Sattler, and M. Truszczynski, editors, *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)*, pages 572 – 574. AAAI Publications, 2010.

[Jackson, 1985] F. Jackson. On the semantics and logic of obligation. *Mind*, 94:177–195, 1985.

[Jago, 2006] M. Jago. *Logics for Resource-Bounded Agents.* PhD thesis, University of Nottingham, 2006.

[Konkka, 2000] J. Konkka. Funk games: Approaching collective rationality. E-Thesis, University of Helsinki, 2000.

[Kooi and Tamminga, 2006] B. Kooi and A. Tamminga. Conflicting obligations in multi-agent deontic logic. In J.-J. Ch. Meyer and L. Goble, editors, *Deontic Logic and Artificial Normative Systems: 8th International Workshop on Deontic Logic in Computer Science*, volume 4048 of *LNCS*, pages 175–186. Springer, 2006.

[Lamarre and Shoham, 1994] P. Lamarre and Y. Shoham. Knowledge, certainty, belief, and conditionalisation abbreviated version. In J. Doyle, E. Sandewall, and P. Torasso, editors, *KR'94*, Morgan Kaufmann 1994. pages 415–424.

[Lamarre, 1991] P. Lamarre. S4 as the conditional logic of nonmonotonicity. In J. F. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, Morgan Kaufmann 1991. pages. 357–367, 1991.

[Lang and van der Torre, 2008] J. Lang and L. van der Torre. From belief change to preference change. In M. Ghallab, C. D. Spyropoulos, N. Fakotakis, and N. Avouris, editors, *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI-2008)*, pages 351–355, 2008.

[Lang *et al.*, 2003] J. Lang, L. van der Torre, and E. Weydert. Hidden uncertainty in the logical representation of desires. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. pages 189–231, 2003.

[Lewis, 1973] D. Lewis. *Counterfactuals*. Oxford: Blackwell, 1973.

[Liu *et al.*, 2014] F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*, 191:2403–2431, 2014.

[Liu, 2004] F. Liu. Dynamic variations: Update and revision for diverse agents. Master's thesis, MoL-2004-05. ILLC, University of Amsterdam, 2004.

[Liu, 2008] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity*. PhD thesis, ILLC, University of Amsterdam, 2008.

[Liu, 2009a] F. Liu. Logics for interaction between preference and belief. Manuscript, Department of Philosophy, Tsinghua University, 2009.

[Liu, 2009b] F Liu. Preference change: A quantitative approach. *Studies in Logic*, 2(3):12–27, 2009.

[Liu, 2011a] F. Liu. *Reasoning about Preference Dynamics*, volume 354 of *Synthese Library*. Springer, 2011.

[Liu, 2011b] F. Liu. A two-level perspective on preference. *Journal of Philosophical Logic*, 40:421–439, 2011.

[Loohuis, 2009] L.O. Loohuis. Obligations in a responsible world. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 251–262. Springer, 2009.

[Mastop, 2005] R. Mastop. *What Can You Do? Imperative Mood in Semantic Theory*. PhD thesis, ILLC, University of Amsterdam, 2005.

[Meyden, 1996] R. van der Meyden. The dynamic logic of permission. *Journal of Logic and Computation*, 6:465–479, 1996.

[Meyer and van der Hoek, 1995] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for Computer Science and Artificial Intelligence*. Cambridge: Cambridge University Press, 1995.

[Meyer, 1988] J-J.Ch. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

[Moore, 1903] G. E. Moore. *Principia Ethica*. Cambridge University Press, 1903.

[Osherson and Weinstein, 2012a] D. Osherson and S. Weinstein. Preference Based on Reasons. *The Review of Symbolic Logic*, 5:122–147, 3 2012.

[Osherson and Weinstein, 2012b] D. Osherson and S. Weinstein. Quantified Preference Logic. arXiv:1208.2921, 2012.

[Osherson and Weinstein, 2014] D. Osherson and S. Weinstein. Deontic modality based on preference. arXiv:1409.0824, 2014.

[Otterloo, 2005] S. van Otterloo. *A Strategic Analysis of Multi-agent Protocols*. PhD thesis, Liverpool University, UK, 2005.

[Pacuit *et al.*, 2006] E. Pacuit, R. Parikh, and E. Cogan. The logic of knowledge based on obligation. *Synthese*, 149:311–341, 2006.

[Parent, 2003] X. Parent. Remedial interchange, contrary-to-duty obligation and commutation. *Journal of Applied Non-Classical Logics*, 13(3/4):345–375, 2003.

[Parent, 2014] X. Parent. Maximality vs. optimality in dyadic deontic logic. *Journal of Philosophical Logic*, 43:1101–1128, 2014.

[Parent, 2015] X. Parent. Completeness of Åqvist's systems E and F. *The Review of Symbolic Logic*, 8:164–177, 2015.

[Parikh and Ramanujam, 2003] R. Parikh and R. Ramanujam. A knowledge-based semantics of messages. *Journal of Logic, Language and Information*, 12:453–467, 2003.

[Parikh, 1999] R. Parikh. Beliefs, belief revision, and splitting languages. In J. Ginzburg, L. Moss, and M. de Rijke, editors, *Logic, Language and Computation*, volume 2, pages 266–278. Center for the Study of Language and Information Stanford, CA, 1999.

[Plaza, 1989] J.A. Plaza. Logics of public communications. In M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216, 1989.

[Prince and Smolensky, 2004] A. Prince and P. Smolensky. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell, 2004.

[Pucella and Weissmann, 2004] R. Pucella and V. Weissmann. Reasoning about dynamic policies. In I. Walukiewicz, editor, *Proceedings FoSSaCS-7*, Lecture Notes in Computer Science 2987, pages 453–467, 2004.

[Rao and Georgeff, 1991] A.S. Rao and M.P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484. San Mateo, CA: Morgan Kaufmann, 1991.

[Rescher, 1966] N. Rescher. Notes on preference, utility, and cost. *Synthese*, 16:332–343, 1966.

[Rott, 2003] H. Rott. Basic entrenchment. *Studia Logica*, 73:257–280, 2003.

[Rott, 2006] H. Rott. Shifting priorities: Simple representations for 27 iterated theory change operators. In H. Langerlund, S. Lindström, and R. Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, pages 359–384. Uppsala Philosophical Studies 53, 2006.

[Roy *et al.*, 2014] O. Roy, A.J. Anglberger, and N. Gratzl. The logic of best actions from a deontic perspective. In A. Baltag and S. Smets, editors, *Johan van Benthem on Logic and Information Dynamics*, page 657–76. Springer, 2014.

[Roy, 2008] O. Roy. *Thinking before Acting: Intentions, Logic and Rational Choice*. PhD thesis, ILLC, University of Amsterdam, 2008.

[Searle and Veken, 1985] J. R. Searle and D. van der Veken. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press, 1985.

[Segerberg, 2001] K. Segerberg. The basic dynamic doxastic logic of AGM. In M.-A. Williams and H. Rott, editors, *Frontiers in Belief Revision*, pages 57–84. Kluwer Academic Publishers, 2001.

[Seligman *et al.*, 2013] J. Seligman, F. Liu, and P. Girard. Facebook and the epistemic logic of friendship. In B. C. Schipper, editor, *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 229–238, 2013.

[Sergot, 2004] M. Sergot. $(C+)^{++}$: An action language for modelling norms and institutions. Technical Report 8, Department of Computing, Imperial College, London, 2004.

[Shi, 2014] C. Shi. Logics of evidence-based belief and knowledge. Master's thesis, Tsinghua University, 2014.

[Shoham and Leyton-Brown, 2008] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations*. Cambridge: Cambridge University Press, 2008.

[Shoham, 1988] Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. Cambridge, MA: The MIT Press, 1988.

[Snyder, 2004] J. Snyder. Product update for agents with bounded memory. Manuscript, Department of Philosophy, Stanford University, 2004.

[Spohn, 1988] W. Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W.L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change and Statistics II*, pages 105–134. Dordrecht: Kluwer, 1988.

[Stalnaker, 1984] R. Stalnaker. *Inquiry*. Cambridge University Press: Cambridge, United Kingdom, 1984.

[Stalnaker, 1996] R. Stalnaker. Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12(2):133–163, 1996.

[Stalnaker, 1999] R. Stalnaker. Extensive and strategic forms: Games and models for games. *Research in Economics*, 53:293–319, 1999.

[Stalnaker, 2006] R. Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128(1):169–199, 2006.

[Tamminga, 2013] A. Tamminga. Deontic logic for strategic games. *Erkenntnis*, 78(1):183–200, 2013.

[Tan and van der Torre, 1996] Y.-H. Tan and L. van der Torre. How to combine ordering and minimizing in a deontic logic based on preferences. In M. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems, DEON '96: The 3rd International Workshop on Deontic Logic in Computer Science*, pages 216–232, 1996.

[Tosatto *et al.*, 2012] S. C. Tosatto, G. Boella, L. van der Torre, and S. Villata. Abstract normative systems: Semantics and proof theory. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*, pages 358–368. AAAI Publications, 2012.

[Trapp, 1985] R.W. Trapp. Utility theory and preference logic. *Erkenntnis*, 22:301–339, 1985.

[van der Torre and Tan, 1999] L. van der Torre and Y.-H. Tan. An update semantics for deontic reasoning. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 73–90. Amsterdam: IOS Press, 1999.

[van der Torre, 1997] L. van der Torre. *Reasoning about Obligations: Defeasibility in Preference-based Deontic Logic*. PhD thesis, Rotterdam, 1997.

[van Ditmarsch *et al.*, 2007] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Berlin: Springer, 2007.

[Velazquez-Quesada, 2009] F. R. Velazquez-Quesada. Inference and update. *Synthese*, 169:283–300, 2009.

[Veltman, 1985] F. Veltman. *Logics for Conditionals*. PhD thesis, University of Amsterdam, 1985.

[Veltman, 1996] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

[von Wright, 1951] G. H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

[von Wright, 1956] G. H. von Wright. A note on deontic logic and derived ogligation. *Mind*, 65:507–509, 1956.

[von Wright, 1963] G. H. von Wright. *The Logic of Preference*. Edinburgh: Edinburgh University Press, 1963.

[von Wright, 1964] G. H. von Wright. A new system of deontic logic. In *Danish Yearbook of Philosophy*, pages 173–182. Museum Tusculanum Press, Copenhagen, 1964.

[Wellman and Doyle, 1991] M. Wellman and J. Doyle. Preferential semantics for goals. In *Proceedings of the National Conference on Artificial Intelligence*, pages 698–703, 1991.

[Wooldridge, 2000] M. Wooldridge. *Reasoning about Rational Agents*. Cambridge, MA: The MIT Press, 2000.

[Yamada, 2006] T. Yamada. Acts of commands and changing obligations. In K. Inoue, K. Satoh, and F. Toni, editors, *Proceedings of the 7th Workshop on Computational Logic in Multi-Agent Systems (CLIMA VII)*, 2006. Revised version appeared in LNAI 4371, pages 1-19, Springer-Verlag, 2007.

[Yamada, 2008] T. Yamada. Logical dynamics of some speech acts that affect obligations and preferences. *Synthese*, 165(2):295–315, 2008.

[Yamada, 2010] T. Yamada. Scorekeeping and dynamic logics of speech acts. Manuscript, Hokkaido University, 2010.

[Zarnic, 2003] B. Zarnic. Imperative change and obligation to do. In K. Segerberg and R. Sliwinski, editors, *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Aqvist*, pages 79–95. Uppsala philosophical studies 51. Uppsala: Department of Philosophy, Uppsala University, 2003.

Johan van Benthem
Institute for Logic, Language and Computation (ILLC),
University of Amsterdam,
The Netherlands.

Changjiang Distingushed Professor, Ministry of Education,
Tsinghua University, Beijing, 100084, China
Email: johan.vanbenthem@uva.nl

Fenrong Liu
Department of Philosophy,
Tsinghua University, Beijing, 100084, China
Email: fenrong@tsinghua.edu.cn

HENRI PRADE AND GILLES RICHARD

HOMOGENEOUS AND HETEROGENEOUS LOGICAL PROPORTIONS: AN INTRODUCTION

1 INTRODUCTION

Commonsense reasoning often relies on the perception of similarity as well as dissimilarity between objects or situations. Such a perception may be expressed and summarized by means of analogical proportions, i.e., statements of the form "A is to B as C is to D". Analogy is not a mere question of similarity between two objects (or situations), but rather a matter of proportion or relation between objects. This view dates back to Aristotle and was enforced by Scholastic philosophy. Indeed, an analogical proportion equates a relation between two objects with the relation between two other objects. As such, the analogical proportion "A is to B as C is to D" poses an analogy of proportionality by (implicitly) stating that the way the two objects A and B, otherwise similar, differ is the same way as the two objects C and D, which are similar in some respects, differ.

A propositional logic modeling of analogical proportions, viewed as a quaternary connective between the Boolean values of some feature pertaining to A, B, C, and D, has been recently proposed in [Miclet and Prade, 2009]. This logical modeling amounts to precisely state that the difference between A and B is the same as the one between C and D, and that the difference between B and A is the same as the one between D and C. This view can then be proved to be equivalent to state that each time a Boolean feature is true for A and D (resp. A or D) it is also true for B and C (resp. B or C), and conversely. This latter point shows that a counterpart of a characteristic behavior of numerical geometrical proportions ($\frac{a}{b} = \frac{c}{d}$), or of numerical arithmetic proportions ($a-b = c-d$), namely that the product (resp. sum, in the second case) of the extremes is equal to the product (resp. the sum) of the means, is still observed in the logical setting.

However, analogical proportions are not the only type of quaternary statements relying on the ideas of similarity and dissimilarity that can be imagined. They turn out to be a special case of so-called *logical proportions* [Prade and Richard, 2010a]. Roughly speaking, a logical proportion between four terms A, B, C, D equates similarity or dissimilarity evaluations about the pair (A, B) with similarity or dissimilarity evaluations about the

D. M. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*,
Handbook of Philosophical Logic 18, https://doi.org/10.1007/978-3-319-97755-3_2

pair (C, D). A set of 120 distinct logical proportions, whose formal expressions share the same structure as well as some remarkable properties, has been identified. Among them, 8 logical proportions stand out as being the only ones that enjoy a code independency property. Namely, their truth status remains unchanged when the truth values 0 and 1 are exchanged. These 8 proportions split into two groups, namely, 4 *homogeneous* ones (which include the analogical proportion) [Prade and Richard, 2012b], and 4 *heterogeneous* logical proportions, which are dual in some sense of the former ones. The pairs (A, B) and (C, D) play symmetrical roles for homogeneous proportions, while it is not the case for the heterogeneous ones. However, both enjoy noticeable permutation properties.

Similarity and dissimilarity are naturally a matter of degrees. Thus, the extension of homogeneous and heterogeneous logical proportions when features are graded make sense in a multiple-valued logic setting. This makes these logical proportions closer to a symbolic counterpart of numerical proportions where the equality between ratios or differences of quantities may be approximate.

Besides, knowing three values, the statement of the equality of numerical ratios, or of numerical differences, involving a fourth unknown value, and expressing a proportionality relation, is useful for extrapolating this latter value. Similarly, the solving of logical proportion equations may be the basis of reasoning procedures. In particular, when an analogical proportion holds for a large number of features between four situations described by means of n binary features, one may make the plausible inference that the same type of proportion should also hold for a $(n + 1)$th feature. If the truth value of this latter feature is known for three of the situations, and unknown for the fourth one, this value can thus be obtained as the solution of an analogical proportion equation.

The paper is organized as follows. In Section 2, the notion of logical proportions is introduced and formally defined. Then, a structural typology of the different families of logical proportions, as well as some noticeable properties, are presented. Section 3 is devoted to a more detailed study of homogeneous proportions. Section 4 deals with extensions of homogeneous proportions for handling non Boolean or unknown features. This is the case if the features are gradual, or if they are binary but may not apply. It may also happen that for some situations it is not known if a feature holds or not. The section investigates these three types of cases (gradual features, features non applicable, and missing information about a feature), where different multiple-valued logical calculi are involved. Section 5 focuses on heterogeneous proportions, studies their properties, and their extension to gradual properties. Section 6 discusses applications of homogeneous and heterogeneous proportions. Homogeneous logical proportions, especially analogical proportions, seem of interest for completing missing values in tables, a problem sometimes termed "matrix abduction " [Abraham *et al.*, 2009]. It amounts in the logical proportion setting to completing a series A, B, C with X such as (A, B, C, X) makes a proportion of a given type. Heterogeneous logical proportions are shown to be instrumental for picking out the item that does not fit in a list. Thus, the setting of logical proportions appears to be rich enough for coping with two different types of reasoning problems where the ideas of similarity and dissimilarity play a key role in both cases. Psychological quizzes or tests are used for illustrating this ability to exploit comparisons in reasoning.

This paper provides a synthesis of results that have appeared mostly in a series of papers by the authors [Prade and Richard, 2010c; Prade and Richard, 2010b; Prade and Richard, 2012b].

2 LOGICAL PROPORTIONS

Before introducing the formal definitions, let us briefly clarify the notations used.

- When dealing with Boolean logic, $a, b, \ldots$ denote propositional variables (having 0 or 1 as truth value), and we use the standard symbols $\wedge, \vee$ to build up formulas (with parentheses when needed). For the negation operator, instead of using the standard $\neg$ symbol, we will use $\overline{a}$ to denote $\neg a$. This is done for saving space when writing long formulas. As usual $\top$ (resp. $\bot$) denotes the always true (resp. false) proposition.

- 0 and 1 denote the Boolean truth values, and a valuation v is just a function from the set of propositional variables to the set of truth values, i.e., $\{0, 1\}$ in the Boolean case, or $[0, 1]$ in the graded case.

- When we propose a new definition, we will use the symbol $\triangleq$ meaning definitional equality. The right hand side of the equation is the definition of the left hand side.

- When we consider syntactic identity, we use $=_{Id}$: for instance $a \wedge b =_{Id} a \wedge b$ but we do not have $a \wedge b =_{Id} b \wedge a$.

- Finally, the symbol $\equiv$ is reserved for the equivalence, i.e.,

$$a \to b \triangleq \overline{a} \vee b \qquad a \equiv b \triangleq (a \to b) \wedge (b \to a)$$

Logical proportions are Boolean formulas built upon what we called indicators. We introduce this concept in the next subsection and we investigate some fundamental properties.

2.1 Similarity and dissimilarity indicators

Generally speaking, the comparison of two items A and B relies on the representation of these items. For instance, the items may be represented as a set of features $\mathcal{A}$ and $\mathcal{B}$. Then, one may define a *similarity measure*. This is the aim of the well-known work of Amos Tversky [1977], taking into account the common features, the specificities of A w.r.t. B, and the specificities of B w.r.t. A, respectively modeled by $\mathcal{A} \cap \mathcal{B}$, $\mathcal{A} \setminus \mathcal{B}$, and $\mathcal{B} \setminus \mathcal{A}$. Here, we are not looking for any global measure of similarity, we are rather interested in keeping track in what respect items are similar and in what respect they are dissimilar using Boolean indicators. This is why we adopt a logical setting: features are viewed as Boolean properties. Let P be such a property, which can be seen as a predicate: $P(A)$ may be true (in that case $\neg P(A)$ is false), or false.

When comparing two items A and B w.r.t. such a property P, it makes sense to consider A and B similar (w.r.t. property P):
- when $P(A) \wedge P(B)$ is true or
- when $\neg P(A) \wedge \neg P(B)$ is true.

In the remaining cases:
- when $\neg P(A) \wedge P(B)$ is true or
- when $P(A) \wedge \neg P(B)$ is true,

we can consider A and B as dissimilar w.r.t. property P.

Since $P(A)$ and $P(B)$ are ground formulas, they can simply be considered as Boolean variables, and denoted a and b by abstracting w.r.t. P. If the conjunction $a \wedge b$ is true, the property is satisfied by both items A and B, while the property is satisfied by neither A nor B if $\overline{a} \wedge \overline{b}$ is true. The property is true for A only (resp. B only) if $a \wedge \overline{b}$ (resp. $\overline{a} \wedge b$) is true. This is why we call such a conjunction of Boolean literals an *indicator*, and for a given pair of Boolean variables (a, b), we have exactly 4 distinct indicators:

- $a \wedge b$ and $\overline{a} \wedge \overline{b}$ that we call *similarity indicators*,
- $a \wedge \overline{b}$ and $\overline{a} \wedge b$ that we call *dissimilarity indicators*.

Let us observe that negating anyone of the two terms of a dissimilarity indicator turns it into a similarity indicator, and conversely. Hence, negating the two terms of an indicator yields an indicator of the same type.

2.2 Building logical proportions with indicators

When describing two elementary situations encoded by two Boolean variables a and b, one may use one of the four above indicators. Putting such a description in relation with what takes place with two other Boolean variables c and d in terms of some indicator, leads to state an equivalence between one indicator pertaining to the pair (a, b) and one indicator pertaining to the pair (c, d). However, one may consider that using *two* indicators to describe the status of 2 variables a and b may be more satisfactory from some symmetrization point of view than using only one indicator. For instance, using $\overline{a} \wedge b$ together with $a \wedge \overline{b}$ establishes the symmetry between a and b, or using $a \wedge \overline{b}$ together with $a \wedge b$ considers counter-examples as well as examples in context a, or using $\overline{a} \wedge \overline{b}$ together with $a \wedge b$ provides the same role to negative or positive features. Note that such symmetrizations occur for free with numerical proportions where for instance one can exchange a and b on the one hand, c and d on the other hand, still writing a unique equality. It is why we more particularly focus on proportions defined as the conjunction of *two distinct* equivalences between an indicator for the pair (a, b) and an indicator for the pair (c, d).

One may wonder about the simultaneous use of three indicators for comparing two Boolean variables. This would lead to three equivalences instead of two, which appears conceptually more complicated, and maybe farther from the idea of proportion inherited from the numerical setting. Then, for the sake of simplicity, we stick to the conjunctions of two equivalences between indicators in the following. This defines a so-called *logical proportion* [Prade and Richard, 2010a; Prade and Richard, 2010c]. More formally, let us denote $I_{(a,b)}$ and $I'_{(a,b)}$[1] (resp. $I_{(c,d)}$ and $I'_{(c,d)}$) 2 indicators for (a, b) (resp. (c, d)). Then

Definition 1. A logical proportion $T(a, b, c, d)$ is the conjunction of 2 distinct equivalences between indicators of the form

$$I_{(a,b)} \equiv I_{(c,d)} \wedge I'_{(a,b)} \equiv I'_{(c,d)}$$

An example of such proportion is $((\overline{a} \wedge \overline{b}) \equiv (c \wedge \overline{d})) \wedge ((\overline{a} \wedge b) \equiv (\overline{c} \wedge d))$ where

- $I_{(a,b)} \triangleq \overline{a} \wedge \overline{b}, \quad I_{(c,d)} \triangleq c \wedge \overline{d},$

[1]Note that $I_{(a,b)}$ (or $I'_{(a,b)}$) refers to one element in the set $\{a \wedge b, \overline{a} \wedge b, a \wedge \overline{b}, \overline{a} \wedge \overline{b}\}$, and should not be considered as a functional symbol. Still, we use this notation for the sake of readability.

- $I'_{(a,b)} \triangleq \overline{a} \wedge b, \quad I'_{(c,d)} \triangleq \overline{c} \wedge d.$

Obviously, this formal definition goes beyond what may be expected from the informal idea of "logical proportion", since equivalences may be put between things that are not homogeneous (i.e., mixing similarity and dissimilarity indicators in various ways).

Let us first determine the number of logical proportions. To build an equivalence between indicators, we have to choose one indicator among four for the pair (a, b) and similarly for the pair (c, d), we get $4 \times 4 = 16$ distinct equivalences. To build up a logical proportion, we first choose one equivalence among 16, and then the second equivalence has to be chosen among the 15 remaining ones, leading to $16 \times 15 = 240$ pairs of equivalences. Taking into account the commutativity of the Boolean conjunction, we finally get $240/2 = 120$ potentially distinct logical proportions . We shall see in subsection 2.4 that they are indeed distinct. We first provide a syntactic typology of the logical proportions.

2.3 Typology of logical proportions

Logical proportions can be classified according to the ways they are built up. At this stage, it makes sense to distinguish between two types of indicators: similarity indicators that are denoted by S, and dissimilarity indicators that are denoted by D: e.g., $D_{(a,b)} \in \{a \wedge \overline{b}, \overline{a} \wedge b\}$.

Depending on the way the indicators are chosen, one may mix the similarity and the dissimilarity indicators differently in the definition of a proportion.

This leads us to distinguish a specific subfamily of proportions, the so-called *degenerated proportions*: those ones involving only 3 distinct indicators in their definition. For instance

$$(a \wedge b \equiv \overline{c} \wedge d) \wedge (\overline{a} \wedge b \equiv \overline{c} \wedge d)$$

is such a proportion where $I_{(c,d)} =_{Id} I'_{(c,d)}$.

For the remaining proportions, it is required that all the indicators appearing in the definition of the proportion are distinct. At this stage, among the non-degenerated proportions, we can identify 4 subfamilies that we describe below:

- **The 4 homogeneous proportions**

  For these proportions, we do not mix different types of indicators in the 2 equivalences. The homogeneous proportions are of the form

$$S_{(a,b)} \equiv S_{(c,d)} \wedge S'_{(a,b)} \equiv S'_{(c,d)}$$

or

$$D_{(a,b)} \equiv D_{(c,d)} \wedge D'_{(a,b)} \equiv D'_{(c,d)}$$

Thus, it appears that only 4 proportions among 120 are homogeneous. They are (with their name):

– *analogy* : $A(a, b, c, d)$, defined by

$$((a \wedge \overline{b}) \equiv (c \wedge \overline{d})) \wedge ((\overline{a} \wedge b) \equiv (\overline{c} \wedge d))$$

– *reverse analogy*: $R(a, b, c, d)$, defined by

$$((a \wedge \overline{b}) \equiv (\overline{c} \wedge d)) \wedge ((\overline{a} \wedge b) \equiv (c \wedge \overline{d}))$$

– *paralogy* : $P(a, b, c, d)$, defined by

$$((a \wedge b) \equiv (c \wedge d)) \wedge ((\overline{a} \wedge \overline{b}) \equiv (\overline{c} \wedge \overline{d}))$$

– *inverse paralogy*: $I(a, b, c, d)$, defined by

$$((a \wedge b) \equiv (\overline{c} \wedge \overline{d})) \wedge ((\overline{a} \wedge \overline{b}) \equiv (c \wedge d))$$

Analogy already appeared under this form in [Miclet and Prade, 2009]; paralogy and reverse analogy were first introduced in [Prade and Richard, 2009], and inverse paralogy in [Prade and Richard, 2010c]. While the analogical proportion (analogy, for short) reads "a is to b as c is to d" and expresses that "a differs from b as c differs from d, and conversely b differs from a as d differs from c", reverse analogy expresses that "a differs from b as d differs from c, and conversely b differs from a as c differs from d", paralogy expresses that "what a and b have in common, c and d have it also" (positively and negatively). Paralogy is a given name. Finally, *inverse paralogy* expresses that "what a and b have in common, c and d miss it, and conversely". As can be seen, inverse paralogy expresses a form of antinomy between pairs (a, b) and (c, d). Note that we use two different words, "inverse" and "reverse", since the changes between analogy and reverse analogy on the one hand, and paralogy and inverse paralogy on the other hand, are not of the same nature. From now on, we denote analogy with A, reverse analogy with R, paralogy with P, inverse analogy with I. When we need to denote any unspecified proportion, we will use the letter T.

- **The 16 conditional proportions**

  Their expression is made of the conjunction of an equivalence between similarity indicators and of an equivalence between dissimilarity indicators. Thus, they are of the form

  $$S_{(a,b)} \equiv S_{(c,d)} \wedge D_{(a,b)} \equiv D_{(c,d)}$$

  There are 16 conditional proportions (2×2 choices *per* equivalence). An example is

  $$((a \wedge b) \equiv (c \wedge d)) \wedge ((a \wedge \overline{b}) \equiv (c \wedge \overline{d}))$$

  Let us explain the term "conditional". It comes from the fact that these proportions express "equivalences" between conditional statements. Indeed, it has been advocated in [Dubois and Prade, 1994] that a rule "if a then b" can be seen as a three valued entity that is called 'conditional object' and denoted $b|a$ [De Finetti, 1936]. This entity is:

  - true if $a \wedge b$ is true. The elements making it true are the examples of the rule "if a then b",
  - false if $a \wedge \overline{b}$ is true. The elements making it true are the counter-examples of the rule "if a then b",
  - undefined if $\overline{a}$ is true. The rule "if a then b" is then not applicable.

  Thus, the above proportion $((a \wedge b) \equiv (c \wedge d)) \wedge ((a \wedge \overline{b}) \equiv (c \wedge \overline{d}))$ may be denoted $b|a :: d|c$ combining the two conditional objects in the spirit of the usual notation for analogical proportion. Indeed, it expresses a semantical equivalence between the 2 rules "if a then b" and "if c then d" by stating that they have the same examples, i.e. $(a \wedge b) \equiv (c \wedge d))$ and the same counter-examples $(a \wedge \overline{b}) \equiv (c \wedge \overline{d})$.

  It is worth noticing that such proportions have equivalent forms, e.g.:

  $$(b|a :: d|c) \equiv (\overline{b}|a :: \overline{d}|c)$$

  which agrees with the above semantics and more generally with the idea of conditioning. Indeed the examples "if a then b" are the counter-examples of "if a then $\overline{b}$", and vice-versa. Due to this remark, it is enough to consider the equivalences between one of the 4 conditional objects $a|b$, $b|a$, $a|\overline{b}$, $b|\overline{a}$, and the 4 other conditional objects built with (c, d), yielding 4×4 proportions as expected. Besides, 8 conditional proportions have been first considered in [Prade and Richard, 2010c], but not the 8 remaining ones, since they do not satisfy the "full identity" property, discussed in the next section.

- **The 20 hybrid proportions**

They are characterized by equivalences between similarity and dissimilarity indicators in their definitions. They are of the form.

$$S_{(a,b)} \equiv D_{(c,d)} \wedge S'_{(a,b)} \equiv D'_{(c,d)}$$

or

$$D_{(a,b)} \equiv S_{(c,d)} \wedge D'_{(a,b)} \equiv S'_{(c,d)}$$

or

$$S_{(a,b)} \equiv D_{(c,d)} \wedge D_{(a,b)} \equiv S_{(c,d)}.$$

There are 20 hybrid proportions: 2 of the first type, 2 of the second type, 16 of the third type since we have here 4 choices for an equivalence $S_{(a,b)} \equiv D_{(c,d)}$, and 4 choices for $D_{(a,b)} \equiv S_{(c,d)}$.

If we remember that negating anyone of the two terms of a dissimilarity indicator turns it into a similarity indicator, and conversely, we understand that changing a into $\overline{a}$ (and $\overline{a}$ into a), or applying a similar transformation with respect to b, c, or d, turns

- an hybrid proportion into an homogeneous or a conditional proportion;

- an homogeneous or a conditional proportion into an hybrid proportion.

This indicates the close relationship of hybrid proportions with homogeneous and conditional proportions. More precisely,

- on the one hand there are 4 hybrid proportions such that replacing a with $\overline{a}$ leads to the 4 homogeneous proportions A, R, P, I. They are obtained by the two first kinds of patterns for building hybrid proportions. Moreover, we shall see in the next section that they constitute with the 4 homogeneous proportions the 8 proportions that are the only ones satisfying "code independency " property.

- on the other hand, there are 16 remaining hybrid proportions, obtained by the third kind of pattern for building them. They can be written as the equivalence of 2 conditional objects, although they do not obey the conditional proportion pattern. For instance, $((\overline{a} \wedge b) \equiv (c \wedge d)) \wedge ((a \wedge b) \equiv (\overline{c} \wedge d))$ can be written as $\overline{a}|b :: c|d$. This proportion is indeed obtained from the conditional proportion $a|b :: c|d$ by changing a into $\overline{a}$. Thus, these 16 new equivalences between conditional objects are not of the form $a|b :: c|d$ (or equivalently $\overline{a}|b :: \overline{c}|d$) produced by the pattern of conditional proportions, but of a "mixed" form having an odd number of negated terms.

- **The 32 semi-hybrid proportions**

  One half of their expressions involve indicators of the same type, while the other half requires equivalence between indicators of opposite types. They are of the form

$$S_{(a,b)} \equiv S_{(c,d)} \wedge S'_{(a,b)} \equiv D_{(c,d)}$$

or

$$S_{(a,b)} \equiv S_{(c,d)} \wedge D_{(a,b)} \equiv S'_{(c,d)}$$

or

$$D_{(a,b)} \equiv D_{(c,d)} \wedge S_{(a,b)} \equiv D'_{(c,d)}$$

or

$$D_{(a,b)} \equiv D_{(c,d)} \wedge D'_{(a,b)} \equiv S_{(c,d)}$$

There are 32 semi-hybrid proportions (8 of each kind: 4 choices for the first equivalence, times 2 choices for the element that is not of the same type as the three others (D or S) in the second equivalence). An example of semi-hybrid proportion is $((a \wedge b) \equiv (c \wedge d)) \wedge ((\overline{a} \wedge \overline{b} \equiv (\overline{c} \wedge d))$.

Applying a change from a to $\overline{a}$ (and $\overline{a}$ to a), or applying a similar transformation with respect to b, c, or d, turns a semi-hybrid proportion into a semi-hybrid proportion (since as already said, negating anyone of the two terms of a dissimilarity indicator turns it into a similarity indicator, and conversely). This contrasts with the hybrid proportion class which is not closed under such a transformation.

- **The 48 degenerated proportions**

  In all the above categories, the 4 indicators related by equivalence symbols should be all distinct. In degenerated proportions, there are only 3 different indicators and it is simpler to come back to our initial notation. With this notation, these proportions are of the form

$$I_{(a,b)} \equiv I_{(c,d)} \wedge I_{(a,b)} \equiv I'_{(c,d)}$$

or

$$I_{(a,b)} \equiv I_{(c,d)} \wedge I'_{(a,b)} \equiv I_{(c,d)}$$

Their number is easy to compute: we have to choose $I_{(a,b)}$ among 4 indicators and then to choose 2 distinct indicators among 4 pertaining to (c,d): we then get $4 * 6 = 24$ proportions of the first form. The same reasoning with the second kind of expression leads to a total of 48 degenerated proportions. Note that the change from a to $\overline{a}$ (and $\overline{a}$ to a), or a similar transformation with respect to b, c, or d, turns a degenerated proportion into a degenerated proportion.

It can be seen that degenerated proportions always involve a mutual exclusiveness condition between 2 positive or negative literals pertaining to either the pair (a,b) or the pair (c,d). Indeed, if we consider the first form, we get $I_{(a,b)} \equiv I_{(c,d)}$ on the one hand, and $I_{(c,d)} \equiv I'_{(c,d)}$ on the other hand, i.e. an equivalence between two syntactically distinct indicators pertaining to the same pair (c,d). There are 6 cases only:

- $(\overline{c} \wedge d) \equiv (c \wedge \overline{d})$ iff $c \equiv d$
- $(c \wedge d) \equiv (\overline{c} \wedge \overline{d})$ iff $c \equiv \overline{d}$
- $(c \wedge d) \equiv (c \wedge \overline{d})$ iff $c \equiv \perp$
- $(c \wedge d) \equiv (\overline{c} \wedge d)$ iff $d \equiv \perp$
- $(\overline{c} \wedge d) \equiv (\overline{c} \wedge \overline{d})$ iff $\overline{c} \equiv \perp$
- $(c \wedge \overline{d}) \equiv (\overline{c} \wedge \overline{d})$ iff $\overline{d} \equiv \perp$

Thus, we also have $I_{(a,b)} \equiv \perp$ (since we have $I_{(c,d)} \equiv \perp$ and $I'_{(c,d)} \equiv \perp$), which expresses a mutual exclusiveness condition. Since we have 4 possible choices for $I_{(a,b)}$, it yields $4 \times 6 = 24$ distinct proportions, and exchanging (a,b) with (c,d) gives the 24 other degenerated proportions. Generally speaking, degenerated proportions correspond to a mutual exclusiveness condition between component(s) or negation of component(s) of one of the pairs (a,b) or (c,d), together with

- either an identity condition pertaining to the other pair,

- or a tautology condition on one of the literals of the other pair without any constraint on the other literal.

2.4 Basic properties of logical proportions

In this subsection, we first establish a remarkable property that single out the logical proportion s among the whole set of quaternary Boolean formulas. In order to do that we need a lemma.

Lemma 1. An equivalence between indicators has exactly 10 valid valuations.

Proof: Such an equivalence $eq \triangleq I_{a,b} \equiv I_{c,d}$ is satisfied only when it matches one of the 2 patterns $1 = 1$ or $0 = 0$: due to the fact that 0 is an absorbing value for $\wedge$, these patterns correspond to the 10 valuations shown in Table 1 for the literals involved in the indicators (with obvious notation). Any other valuation[2] does not match anyone of the 2 previous patterns and will lead to the truth value 0 for the equivalence eq. □

Table 1. 10 valid valuations for an equivalence between indicators

literal 1	*literal 2*	*literal 3*	*literal 4*	*pattern*
1	1	1	1	$1 = 1$
0	1	0	1	$0 = 0$
0	1	1	0	$0 = 0$
0	1	0	0	$0 = 0$
1	0	0	1	$0 = 0$
1	0	1	0	$0 = 0$
1	0	0	0	$0 = 0$
0	0	0	1	$0 = 0$
0	0	1	0	$0 = 0$
0	0	0	0	$0 = 0$

Proposition 1. The truth table of a logical proportion has 6 and only 6 valuations with truth value 1.

Proof: Since a logical proportion T is the conjunction $eq_1 \wedge eq_2$ of 2 equalities between indicators, with $eq_1 \neq eq_2$, it appears from Lemma 1 that T has a maximum of 10 valid valuations and a minimum of 4 valid valuations. Let us start from $eq1$, having 10 valid valuations which are candidate to validate T. Obviously, adding eq_2 to eq_1 will reduce the number of valid valuations for T. Let us assume eq_2 differs from eq_1 with only one literal (or negation operator). This is then a degenerated proportion. Without loss of

[2]The only valuations considered in this paper pertain to 4-tuples of variables. In practice, a Boolean valuation v will be denoted by the values $v(a)v(b)v(c)v(d)$ without any blank space, e.g., 0100 is short for $v(a) = 0, v(b) = 1, v(c) = 0, v(d) = 0$.

generality, we can consider that the difference between eq_1 and eq_2 occurs on the first literal meaning eq_1 is $a \wedge l_2 \equiv l_3 \wedge l_4$ and eq_2 is $\overline{a} \wedge l_2 \equiv l_3 \wedge l_4$ or vice versa. It is then quite clear that the first valuation 1111 valid for eq_1 is not valid any more for T. It remains 9 candidates valuations. Finally any valuation starting with 01 is not valid any more and we have 3 such valuations. All the 6 remaining valuations are still valid for T. Which ends the proof when the 2 equalities differ from one negation (i.e. one literal). Now when they differ from 2 literals, two cases have to be considered:

- either the 2 literals where eq_1 differs from eq_2 are on the same side of an equivalence i.e. eq_2 is $l_1' \wedge l_2' \equiv l_3 \wedge l_4$ (degenerated proportion)

- or they are on different side i.e. eq_2 is $l_1' \wedge l_2 \equiv l_3' \wedge l_4$.

In the first case, the valuations $1111, 0010, 0001$ and 0000 are not valid any more, but all other ones remain valid. In the second case, the valuations $0100, 0110, 1001$ and 0001 are not valid anymore, but all the other ones remain valid. We are done for the case of 2 differences. When they differ from 3 literals, let us suppose l_4 appears in both equivalence, the valuations $1001, 0101, 0010$ and 0000 are not valid anymore and we stick with the 6 remaining ones. In the case where all the literals are different, obviously the 4 valuations containing only one occurrence of 1 are not valid anymore because they lead to an invalid pattern 0=1 or 1=0 for eq_2. And we have exactly 4 such valuations. It remains 6 valid valuations. $\square$

Note that the negation of a logical proportion is not a logical proportion since such a negation has 10 valuations leading to true in its table. Besides, the 120 logical proportions are all distinct as shown below with the help of the following lemma.

Lemma 2. Two equivalences between indicators have the same truth table iff they are identical.

Proof: It is sufficient to show that if 2 equalities eq_1 and eq_2 have the same truth table, then they are syntactically identical. In other terms, we have to prove that $eq_1 \equiv eq_2$ implies $eq_1 =_{Id} eq_2$. Without loss of generality, let us assume that eq_1 contains a but eq_2 contains $\overline{a}$. Considering the unique valuation v such that $v(eq_1) = 1$ with the pattern $1 = 1$, v is such that $v(a) = 1$. By hypothesis, $v(eq_2) = 1$ but in that case with the pattern $0 = 0$ since $v(\overline{a}) = 0$. Let us now modify v into v' such that $v'(a) = v(a) = 0, v'(c) = v(c), v'(d) = v(d)$ and $v'(b) = v(b)$. Obviously v' does not validate eq_1 but validates eq_2 which contradicts the hypothesis. $\square$

Proposition 2. The truth tables of the 120 proportions are all distinct.

Proof: We are going to show that, when 2 proportions $T \triangleq eq_1 \wedge eq_2$ and $T' \triangleq eq'_1 \wedge eq'_2$ have the same truth table, they are syntactically identical (up to a permutation of the 2 equalities). In other words, $T \equiv T'$ implies $T =_{Id} T'$. Starting from $T \equiv T'$, it amounts to show that if eq_1 is syntactically different from eq'_1, eq_1 is syntactically equal to eq'_2. This will complete the proof as a similar reasoning will show that eq_2 is, in the same context, syntactically equal to eq'_1.

In fact, if eq_1 is syntactically different from eq'_1, we can assume for instance without loss of generality that eq_1 contains a but eq'_1 contains $\overline{a}$. Let us consider the unique valuation σ, validating T and T', such that $\sigma(eq_1) = 1$ with the pattern $1 = 1$. Necessarily, this valuation σ is such that $\sigma(a) = 1$. By hypothesis, $\sigma(eq'_1) = 1$ but in that case with the pattern $0 = 0$ since $\sigma(\overline{a}) = 0$. Let us now modify σ into σ' such that $\sigma'(a) = \sigma(a) = 0, \sigma'(c) = \sigma(c), \sigma'(d) = \sigma(d)$ and $\sigma'(b) = \sigma(b)$. Obviously $\sigma'(T) = \sigma'(eq_1) = 0$ but $\sigma'(eq_1) = 1$ still following the pattern $0 = 0$. The only option for having $\sigma(T) = \sigma(T') = 0$ is thus to have $\sigma'(eq'_2) = 0$ which means a belongs to eq'_2. Continuing the same reasoning, we show that $eq_1 =_{Id} eq'_2$ and we infer that if $eq_1 \neq eq'_1$, necessarily $eq_1 =_{Id} eq_2$. $\square$

Combined with the fact that there are $C_{16}^{6} = 8008$ truth tables with 16 lines, this result makes logical proportions quite rare in the world of quaternary Boolean formulas.

An exhaustive investigation of the whole set of logical proportions with respect to various other properties has been done in [Prade and Richard, 2010c; Prade and Richard, 2012b; Prade and Richard, 2012a]. In the next subsection, we focus on one of these properties which allows us to characterize a small subset of remarkable proportions.

2.5 Code independency

Just as a numerical proportion holds independently of the base used for encoding numbers, or of the system of units representing the quantities at hand, it seems desirable that a logical proportion should be independent of the way we encode items in terms of the truth or the falsity of features. It means that the formula defining a proportion T should be valid when we switch 0 to 1 and 1 to 0. The formal expression of this property, that we call *code independency*, writes:

$$T(a,b,c,d) \to T(\overline{a},\overline{b},\overline{c},\overline{d})$$

Surprisingly, this property highlights the fact once more that a single equivalence would not lead to a satisfactory definition for a logical proportion. Indeed, a unique equivalence between indicators, denoted $l_1 \wedge l_2 \equiv l_3 \wedge l_4$, where the l_i's are literals does not satisfy *code independency*, as explained now. If we consider a valuation v such that $v(l_1) = v(l_2) = v(l_3) = 0$ and $v(l_4) = 1$, obviously v makes the equivalence valid since $v(l_1 \wedge l_2) = v(l_3 \wedge l_4) = 0$. But when we switch 0 to 1 and 1 to 0, it appears that the new valuation v' such that $v'(l_1) = v'(l_2) = v'(l_3) = 1$ and $v'(l_4) = 0$ does not validate the equivalence anymore. This shows that one equivalence is not enough if we are interested in "code independency". We have to consider at least 2 equivalences to capture this behavior. For instance, $(a \wedge b \equiv c \wedge d) \wedge (\overline{a} \wedge \overline{b} \equiv \overline{c} \wedge \overline{d})$ clearly satisfies code independency.

Unfortunately, being built as the conjunction of two equivalences is not a sufficient condition for code independency, and many logical proportions do not satisfy it. We have the following result:

Proposition 3. There are exactly 8 proportions satisfying the code independency property: the 4 homogeneous proportions A, R, P, I, and 4 hybrid proportions (shown in Table 2).

Proof: In fact, the code independency property implies a complete equivalence:

$$T(a,b,c,d) \leftrightarrow T(\overline{a},\overline{b},\overline{c},\overline{d})$$

Since both $T(a,b,c,d)$ and $T(\overline{a},\overline{b},\overline{c},\overline{d})$ are logical proportions, Proposition 2 tells us that the 2 proportions should be identical up to a permutation of the 2 equalities. This exactly means that the second equivalence is obtained from the first one by negating all the variables. Since we have 4×4 equalities between indicators, we can build exactly $16/2 = 8$ proportions satisfying code independency property: each time we choose an equivalence, we use it and its negated form to build up a suitable proportion. Since A, R, P, I are built this way, they satisfy code independency. $\square$

As a consequence of this result, this set of 8 proportions stand out of the whole set of 120 proportions. This set of proportions is clearly divided in 2 subsets: the 4 homogeneous proportions on one hand, and the 4 remaining ones, that we call *heterogeneous* proportions, on the other hand. In the next two sections, we first investigate the 4 homogeneous proportions through the angle of a list of meaningful properties, as well as their interrelationships,

Table 2. The 4 hybrid proportions satisfying code independency

$\mathbf{H_a}$	$\mathbf{H_b}$
$(\overline{a} \wedge b \equiv \overline{c} \wedge \overline{d}) \wedge (a \wedge \overline{b} \equiv c \wedge d)$	$(\overline{a} \wedge b \equiv c \wedge d) \wedge (a \wedge \overline{b} \equiv \overline{c} \wedge \overline{d})$
$\mathbf{H_c}$	$\mathbf{H_d}$
$(\overline{a} \wedge \overline{b} \equiv \overline{c} \wedge d) \wedge (a \wedge b \equiv c \wedge \overline{d})$	$(\overline{a} \wedge \overline{b} \equiv c \wedge \overline{d}) \wedge (a \wedge b \equiv \overline{c} \wedge d)$

and their extensions to multiple-valued settings. After which, we shall move to the study of the 4 heterogeneous proportions in Section 5.

3 THE 4 HOMOGENEOUS PROPORTIONS

We investigate now the 4 homogeneous proportions A, R, P, I from a semantical point of view. When considered as Boolean formulas, their semantics is given via their truth tables (which have $2^4 = 16$ lines since these proportions involve 4 variables).

3.1 Boolean truth tables

Starting from their syntactic expressions, it is an easy game to build up the truth tables of proportions A, R, P, I: they are exhibited in Table 3, where only the valuations leading to the truth value 1, are shown. This means that all the other ones lead to the truth value 0. As expected, only 6 valuations among 16 in the tables lead to a truth value 1. We also observe that there are only 8 distinct valuations that appear in Table 3. This emphasizes their collective coherence as the whole class of homogeneous proportions. Moreover, they go by pairs where 0 and 1 are exchanged, thus pointing out their "code independency".

Table 3. Analogy, Reverse analogy, Paralogy, Inverse paralogy truth tables

A				**R**				**P**				**I**			
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
0	0	1	1	0	0	1	1	1	0	0	1	1	0	0	1
1	1	0	0	1	1	0	0	0	1	1	0	0	1	1	0
0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1
1	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0

It is interesting to take a closer look at the truth tables of the four homogeneous proportions. First, one can observe in Table 3, that 8 possible valuations for (a, b, c, d) never appear among the patterns that make A, R, P, or I true: these 8 valuations are of the form $xxxy$, $xxyx$, $xyxx$, or $yxxx$ with $x \neq y$ and $(x, y) \in \{0, 1\}^2$. As can be seen, it corresponds to situations where $a = b$ and $c \neq d$, or $a \neq b$ and $c = d$, i.e., similarity holds between the components of one of the pairs, and dissimilarity holds in the other pair. Moreover, the truth table of each of the four homogeneous proportions, is built in the same manner:

1. 2 lines of the table correspond to the characteristic pattern of the proportion; namely the two lines where one of the two equivalences in its definition holds true under the form $1 \equiv 1$ (rather than $0 \equiv 0$). Thus,
   - A is characterized by the pattern $xyxy$ (corresponding to valuations 1010 and 0101), i.e. we have the same difference between a and b as between c and d;
   - R is characterized by the pattern $xyyx$ (corresponding to valuations 1001 and 0110), i.e., the differences between a and b and between c and d are in opposite directions;
   - P is characterized by the pattern $xxxx$ (corresponding to valuations 1111 and 0000), i.e., what a and b have in common, c and d have it also;
   - I is characterized by the pattern $xxyy$ (corresponding to valuations 1100 and 0011), i.e. what a and b have in common, c and d do not have it, and conversely.
2. the 4 other lines of the truth table of an homogeneous proportion T are generated by the characteristic patterns of the two other proportions that are not opposed to T (in the sense that A and R are opposed, as well as P and I). For these four lines, the proportion holds true since its expression reduces to $(0 \equiv 0) \wedge (0 \equiv 0)$.

Thus, the six lines of the truth table of A that makes it true are induced by the characteristic patterns of A, P, and I[3], the six valuations that makes P true are induced by the characteristic patterns of P, A, and R, and so on for R and I.

[3]The measure of analogical dissimilarity introduced in [Miclet *et al.*, 2008] is 0 for the valuations corresponding to the characteristic patterns of A, P, and I, maximal for the valuations corresponding to the characteristic patterns of R, and takes the same intermediary value for the 8 valuations characterized by one of the patterns $xxxy$, $xxyx$, $xyxx$, or $yxxx$.

3.2 Relevant properties

Before going deeper in the investigation, remember that the Boolean analogical proportion is supposed to be, in a Boolean setting, the counterpart of the classical numerical proportions. Then, it is interesting to consider Boolean counterparts of the properties satisfied by the numerical proportions, other than *code independency.* We list these properties below (with T denoting a logical proportion).

- *Full identity*: A numerical proportion holds when all the numbers are equal, i.e., $a = b = c = d$, which logically translates into

$$T(a, a, a, a)$$

- *Reflexivity*: A numerical proportion holds between (a, b) and (a, b) which logically translates into

$$T(a, b, a, b)$$

  Obviously, *reflexivity* entails *full identity.*

- *Sameness*: A numerical proportion holds between (a, a) and (b, b), which logically translates into

$$T(a, a, b, b)$$

  Still, *sameness* entails *full identity.*

- *symmetry* : We can exchange the pair (a, b) with the pair (c, d) in the numerical proportion, which logically translates into

$$T(a, b, c, d) \to T(c, d, a, b)$$

- *Central (or extreme) permutation* : This is a well known property of numerical proportions, which logically translates into

$$T(a, b, c, d) \to T(a, c, b, d) \text{ (central permutation)}$$

  or

$$T(a, b, c, d) \to T(d, b, c, a) \text{ (extreme permutation)}$$

- *Transitivity*: This property that holds for numerical proportions is logically stated as follows

$$T(a, b, c, d) \wedge T(c, d, e, f) \to T(a, b, e, f)$$

- *Exchange-mirroring*: The negation operator can play for Boolean values the role of an inverse operator for numbers. A numerical proportion holds between a pair (a, b) and the pair(b^{-1}, a^{-1}), which logically translates into

$$T(a, b, \overline{b}, \overline{a})$$

- *Semi-mirroring*: Similarly it is worth to consider

$$T(a, b, \overline{a}, \overline{b})$$

  This property is not satisfied by numerical proportions.

- *Negation-compatibility*: Similarly it is worth to consider

$$T(a, \overline{a}, b, \overline{b})$$

  This property is also not satisfied by numerical proportions.

Investigating the homogeneous proportions with regard to the properties listed above can simply be done with an examination of the truth table of the target proportion. We summarize in Table 4 all the properties satisfied by A, R, P, I: the third column enumerates the homogeneous proportions satisfying the property, respectively named and described in the 1st and 2nd columns.

Table 4. Boolean properties of A, R, P, I

Property name	Formal definition	Proportion
full identity	$T(a, a, a, a)$	A,R,P
reflexivity	$T(a, b, a, b)$	A,P
reverse reflexivity	$T(a, b, b, a)$	R,P
sameness	$T(a, a, b, b)$	A,R
symmetry	$T(a, b, c, d) \to T(c, d, a, b)$	A,R,P,I
permutation of means	$T(a, b, c, d) \to T(a, c, b, d)$	A,I
permutation of extremes	$T(a, b, c, d) \to T(d, b, c, a)$	A,I
all permutations	$\forall i, j, T(a, b, c, d) \to T(p_{i,j}(a, b, c, d))$	I
transitivity	$T(a, b, c, d) \wedge T(c, d, e, f) \to T(a, b, e, f)$	A,P
semi-mirroring	$T(a, b, \overline{a}, \overline{b})$	R,I
exchange mirroring	$T(a, b, \overline{b}, \overline{a})$	A,I
negation compatib.	$T(a, \overline{a}, b, \overline{b})$	P,I

Note that the 4 homogeneous proportions satisfy symmetry: $T(a,b,c,d) = T(c,d,a,b)$, as well as many other properties. In particular, analogical proportion A enjoys properties that parallel properties of numerical proportions: full identity, reflexivity, symmetry, central and extreme permutations, and transitivity.

One can also establish properties linking the homogeneous proportions, which are easily deducible from their definitions in terms of indicators.

Proposition 4.
$A(a,b,c,d) \equiv R(a,b,d,c)$; $A(a,b,c,d) \equiv P(a,d,c,b)$; $A(a,b,c,d) \equiv I(\overline{a},d,\overline{c},b)$

As can be seen, homogeneous proportions are strongly linked together. Especially A, R, P are exchanged through simple permutation s; in that respect, I stands apart. Besides, A, R, P, I are mutually exclusive, as a simple examination of their truth tables reveals that their intersection is empty.

Proposition 5. $A(a,b,c,d) \wedge R(a,b,c,d) \wedge P(a,b,c,d) \wedge I(a,b,c,d) = \bot$

Lastly, having a closer look on the homogeneous proportions, we can easily build Table 5 which gives what $T(a,b,c,d) \wedge T(c,d,e,f)$ entails for the 4 homogeneous proportions.

Table 5. Chaining properties for A, R, P, I

chaining	result	transitivity
$A \wedge A$	A	yes
$R \wedge R$	A	no
$P \wedge P$	P	yes
$I \wedge I$	P	no
$A \wedge R$	R	
$P \wedge I$	I	

All these common properties explain why the homogeneous proportions stand out from the whole set of 120 logical proportions. It makes homogeneous proportions a worth considering Boolean counterpart of numerical proportions.

3.3 Characterization of homogeneous proportions by properties

Some subsets of the properties listed above are sufficient for characterizing one or more homogeneous proportions as unique among the 120 logical proportions. Let us start with the following result:

Proposition 6.

- A, R, P are the unique proportions to satisfy *full identity* and *code independency*.
- A is the only proportion to satisfy *sameness* ($T(a, a, b, b)$) and *reflexivity* ($T(a, b, a, b)$).
- R is the only proportion to satisfy *sameness* and *reverse reflexivity* $T(a, b, b, a)$.
- P is the only proportion to satisfy *reflexivity* and *reverse reflexivity*.
- There is no proportion simultaneously satisfying *sameness*, *reflexivity*, and *reverse reflexivity*.

Proof: The first statement comes from Proposition 3 giving the 8 proportions satisfying code independency, along with an immediate checking of the proportions syntactic form. For instance, H_a defined as $(a \wedge b \equiv c \wedge \overline{d}) \wedge (\overline{a} \wedge \overline{b} \equiv \overline{c} \wedge d)$ is definitely not valid for valuation 0000. The same reasoning applies to all the proportions other than A, R, P.
This is an easy proof for the first 3 following statements since each property generates a set of 4 valid valuations (and two of them yield 6 valid valuations). For instance, *sameness* ($T(a, a, b, b)$ implies that valuations $1111, 0000, 0011, 1100$ should be valid and *reflexivity* ($T(a, b, a, b)$) implies that valuations $1111, 0000, 0101, 1010$, which is the truth table of A.
Let us consider the last statement, having the simultaneous satisfaction of the 3 properties leads to a truth table where the 8 valuations 0000, 1111, 1010, 0101, 0110, 1001, 0011, 1100 are valid: then this cannot be the truth table of a logical proportion. □

It is well known that a valid numerical proportion still holds when we exchange the extreme elements or the mean elements. And we have seen that A and I satisfy both of these permutations. In fact, there are 6 pairwise permutations of the 4 variables appearing in a proportion. So, the behavior of logical proportions w.r.t. these permutations is worth investigating. We denote the permutation of element i and j by $p_{i,j}$: for instance $p_{2,3}$ is the mean permutation while $p_{1,4}$ is the extreme permutation. We can establish the following result:

Proposition 7.

- A is the only proportion to satisfy *reflexivity* and to be stable for $p_{1,4}$ (or $p_{2,3}$).

- A is the only proportion to satisfy *sameness* and to be stable for $p_{1,4}$ (or $p_{2,3}$).
- R is the only proportion to satisfy *sameness* and to be stable for $p_{1,3}$ (or $p_{2,4}$).
- R is the only proportion to satisfy *reverse reflexivity* and to be stable for $p_{1,3}$ (or $p_{2,4}$).
- P is the only proportion to satisfy *reflexivity* and to be stable for $p_{1,2}$ (or $p_{3,4}$).
- P is the only proportion to satisfy *reverse reflexivity* and to be stable for $p_{1,2}$ (or $p_{3,4}$).
- A and I are the only proportions to satisfy *symmetry* and to be stable for $p_{1,4}$ (or $p_{2,3}$).
- P and I are the only proportions to satisfy *symmetry* and to be stable for $p_{1,2}$ (or $p_{3,4}$).
- I is the *unique* logical proportion to satisfy the 6 permutations.

Proof: The proofs are quite similar for the 8 first statements. Let us give an example for the first statement. *reflexivity* means that valuations $0000, 1111, 0011, 1100$ have to be valid. Adding stability for $p_{2,3}$ leads to add 0101 and 1010 as valid valuations. This is the truth table of A.
Let us consider the last statement which is a bit more tricky. It is easy to check that these permutations induce a partition of the set of valuations into 5 classes, each of them being closed for these 6 permutations:

- the class $\{0000\}$ and the class $\{1111\}$
- the class $\{0111, 1011, 1101, 1110\}$
- the class $\{1000, 0100, 0010, 0001\}$
- the class $\{0101, 1100, 0011, 1010, 1001, 0110\}$

Taking into account that a logical proportion is true for only 6 valuations (Proposition 1), we only have 3 options:
- a proportion valid for $\{0000\}$, $\{1111\}$ and $\{0111, 1011, 1101, 1110\}$,
- or for $\{0000\}$, $\{1111\}$ and $\{1000, 0100, 0010, 0001\}$,
- or for $\{0101, 1100, 0011, 1010, 1001, 0110\}$.

It appears that the latter class is just the truth table of inverse paralogy. Lemma 3 that we shall prove below allows us to complete the proof. □

Lemma 3. A logical proportion cannot satisfies the class of valuation $\{0111, 1011, 1101, 1110\}$ or the class $\{1000, 0100, 0010, 0001\}$.

Proof: It is enough to show that this is the case for an equivalence between indicators. So let us consider such an equivalence $l_1 \wedge l_2 \equiv l_3 \wedge l_4$. If this equivalence is valid for $\{0111, 1011\}$, it means that its truth value does not change when we switch the truth value of the 2 first literals from 0 to 1: there are only 2 indicators for a and b satisfying this requirement: $a \wedge b$ and $\overline{a} \wedge \overline{b}$. On top of that, if this equivalence is still valid for $\{1101, 1110\}$, it means that its truth value does not change when we switch the truth value of the 2 last literals from 0 to 1: there are only 2 indicators for c and d satisfying this requirement: $c \wedge d$ and $\overline{c} \wedge \overline{d}$. Then the equivalence $l_1 \wedge l_2 \equiv l_3 \wedge l_4$ is just $a \wedge b \equiv c \wedge d$, $a \wedge b \equiv \overline{c} \wedge \overline{d}$, $a \wedge b \equiv \overline{c} \wedge \overline{d}$ or $\overline{a} \wedge \overline{b} \equiv \overline{c} \wedge \overline{d}$. None of these equivalences satisfies the whole class $\{0111, 1011, 1101, 1110\}$. The same reasoning applies for the other class. □

We summarize the results of this subsection by a pair of properties characterizing a subset of homogeneous proportions, in Table 6 and Table 7. An empty cell means that the corresponding properties do not characterize any subset of homogeneous proportion. For instance, the diagonal cells are all empty because an homogeneous proportion cannot be characterized with only one property.

Table 6. Characteristic properties of A, R, P, I

	full identity	code indep.	symmetry	sameness	reflexivity	rev. reflexivity
full identity		A, R, P				
code indep.	A, R, P		A, R, P, I	A, R	A, P	R, P
symmetry		A, R, P, I		A, R		R, P
sameness		A, R	A, R		A	R
reflexivity		A, P		A		P
rev. reflexivity		R, P	R, P	R	P	

To conclude this section, we establish a result which shows how singular I is among the set of homogeneous proportions.

Proposition 8.

- A logical proportion satisfying 2 properties among *semi-mirroring*,

Table 7. Characteristic properties of A, R, P, I w.r.t. permutations

	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$	$p_{2,3}$	$p_{2,4}$	$p_{3,4}$
sameness		R	A	A	R	
reflexivity	P		A	A		P
rev. reflexivity	P	R			R	P
symmetry	P, I		A, I	A, I		P, I

negation-compatibility and *exchange-mirroring* satisfies the remaining one, and is unique. This is the inverse paralogy I.

- A logical proportion stable for 4 permutations is stable for the 2 remaining ones and is unique. This is the inverse paralogy I.

Proof: Considering the first statement, let us choose for instance *semi-mirroring* and *negation-compatibility*. First of all, we can observe that, for a proportion T to satisfy *semi-mirroring*, means the 4 valuations 1010, 1001, 0110, 0101 are valid. For *negation-compatibility* to be satisfied, the 4 valuations 1100, 0011, 1001, 0110 should be valid. Then the truth table of a proportion satisfying both properties should contains all these valuations i.e. 1010, 1001, 0110, 0101, 1100, 0011. Thanks to Proposition 1, this is the truth table of inverse paralogy I. A similar reasoning applies for the other cases. Regarding the second statement, let us consider a proportion stable for 4 pairwise permutations: since such pairwise permutations generate the full group of permutations of 4 elements, it means this proportion is stable for any permutations. We can consider 2 cases:

- either such a proportion is valid for a valuation having an even number of 0 and other than 0000 and 1111. We can consider this is 0110 for instance. The stability leads to have $0011, 0110, 0101, 1001, 1010$ valid as well: this is the truth table fo i.

- or such a proportion does not have a valid valuation with an even number of 0 other than 0000 and 1111. It means there is a valid valuation with an odd number of 0 like 1000. In that case, the stability w.r.t. the permutations leads to have $1000, 0100, 0010, 0001$ as valid valuations, which is not possible thanks to Lemma 3. □

3.4 Equation solving

The idea of proportion is closely related to the idea of extrapolation, i.e. to guess/compute a new value on the ground of existing values. In the case of geometrical proportions, this leads to the well known "rule of three" where, knowing that $\frac{a}{b} = \frac{c}{x}$ holds, allows us to compute the value of x from a, b, c. In the Boolean setting, if for some reason it is believed or known that a logical proportion holds between 4 binary items, 3 of them being known, then one may try to infer the value of the 4th one, at least when this extrapolation leads to a unique value. For a proportion T, there are exactly 6 valuations v such that:

$$v(T(a,b,c,d)) = 1$$

In our context, the problem can be stated as follows. Given a logical proportion T and a valuation v such that $v(a), v(b), v(c)$ are known, does it exist a Boolean value x such that $v(T(a,b,c,d)) = 1$ when $v(d) = x$, and in that case, is this value unique?

We will refer to this problem as "the equation solving problem", and for the sake of simplicity, a propositional variable a is denoted as its truth value $v(a)$, and we use the equational notation $T(a,b,c,x) = 1$, where $x \in \{0,1\}$ is unknown. First of all, it is easy to see that there are always cases where the equation has no solution. Indeed, the triple a, b, c may take $2^3 = 8$ values, while any proportion T is true only for 6 distinct valuations, leaving at least 2 cases with no solution. For instance, when we deal with analogy A, the equations $A(1,0,0,x)$ and $A(0,1,1,x)$ have no solution. We have the following results

Proposition 9.
The analogical equation $A(a,b,c,x)$ is solvable iff $(a \equiv b) \vee (a \equiv c)$ holds. In that case, the unique solution is $x = a \equiv (b \equiv c)$.
The reverse analogical equation $R(a,b,c,x)$ is solvable iff $(b \equiv a) \vee (b \equiv c)$ holds. In that case, the unique solution is $x = b \equiv (a \equiv c)$.
The paralogical equation $P(a,b,c,x)$ is solvable iff $(c \equiv b) \vee (c \equiv a)$ holds.
In each of the three above cases, *when it exists*, the unique solution is given by $x = c \equiv (a \equiv b)$, i.e. $x = a \equiv b \equiv c$.
The inverse paralogical equation $I(a,b,c,x)$ is solvable iff $(a \not\equiv b) \vee (b \not\equiv c)$ holds. In that case, the unique solution is $x = c \not\equiv (a \not\equiv b)$.

Proof: By immediate investigation of the truth tables. □

The anthropologist, linguist and computer scientist Sheldon Klein [1982; 1983] was the first to propose to solve analogical equations of the form $A(a,b,c,x) = 1$, where x is unknown, as $x = c \equiv (a \equiv b)$, without however providing an explicit definition for $A(a,b,c,d)$, nor distinguishing between A, R, and P. As we can see, the first 3 homogeneous proportions A, R, P

behave similarly. Still, their conditions of equation solvability differ. Moreover, it can be checked that at least 2 of these proportions are always simultaneously solvable. Besides, when they are solvable, there is a common expression that yields the solution.

3.5 Alternative writings for homogeneous proportions

When sticking to the Boolean setting, we can use standard equivalences to get alternative writings for A, R, P, I. First of all, using the De Morgan's laws and the fact that $p \equiv q$ is equivalent to $\overline{p} \equiv \overline{q}$, we get definitions where the internal $\wedge$ are replaced with $\vee$ as shown in Table 8. It means that, in a Boolean setting, indicators involving $\vee$ are a perfect replacement for indicators using $\wedge$.

Table 8. A, R, P, I definitions with $\vee$ operator

A	R
$(a \vee \overline{b} \equiv c \vee \overline{d}) \wedge (\overline{a} \vee b \equiv \overline{c} \vee d)$	$(a \vee \overline{b} \equiv \overline{c} \vee d) \wedge (a \vee \overline{b} \equiv c \vee \overline{d})$
P	I
$(a \vee b \equiv c \vee d) \wedge (\overline{a} \vee \overline{b} \equiv \overline{c} \vee \overline{d})$	$(a \vee b \equiv \overline{c} \vee \overline{d}) \wedge (\overline{a} \vee \overline{b} \equiv c \vee d)$

A more interesting option is to start from the definition of P with indicators

$$(a \wedge b \equiv c \wedge d) \wedge (\overline{a} \wedge \overline{b} \equiv \overline{c} \wedge \overline{d}) \ (P)$$

and to use again De Morgan's laws to rewrite the second equivalence. This leads to a definition of P without any negation that we denote P^*:

$$(a \wedge b \equiv c \wedge d) \wedge (a \vee b \equiv c \vee d) \ (P^*)$$

Then, considering the link between A and P established in Proposition 4, namely $A(a, b, c, d) \equiv P(a, d, c, b)$, it comes another definition for A, without any negation operator:

$$(a \wedge d \equiv b \wedge c) \wedge (a \vee d \equiv b \vee c) \ (A^*)$$

It is noticeable that this latter new definition exactly corresponds to what the psychologist Jean Piaget [1953], called *logical proportion*! However, strangely enough, he has not developed their study nor pointed out their link with analogy.

Thus, since a and d are the extreme variables, b and c the mean variables, the analogical proportion $A(a, b, c, d)$ can be read as "the conjunction (resp. disjunction) of the extremes is equivalent to the conjunction (resp. disjunction) of the means".

Considering the link between A, R, P, I coming from Proposition 4, we can finally get alternative writing denoted A^*, R^*, P^* and I^* that are shown in Table 9.

Table 9. A^*, R^*, P^*, I^* definitions

A^*	R^*
$(a \wedge d \equiv b \wedge c) \wedge (a \vee d \equiv b \vee c)$	$(a \wedge c \equiv b \wedge d) \wedge (a \vee c \equiv b \vee d)$
P^*	I^*
$(a \wedge b \equiv c \wedge d) \wedge (a \vee b \equiv c \vee d)$	$(a \wedge b \equiv \overline{c} \wedge \overline{d}) \wedge (a \vee b \equiv \overline{c} \vee \overline{d})$

Since, in the Boolean setting, the equivalence $T(a, b, c, d) \equiv T^*(a, b, c, d)$ holds (where T denotes any homogeneous proportion among A, R, P, I), one could consider T^* as an alternative writing for T. It is interesting to note that this approach leads to rewrite A, R, P without any negation. We have to be aware that these equivalences, leading to alternative writings, are not necessarily valid outside the Boolean framework.

4 HOMOGENEOUS PROPORTIONS: MULTIPLE-VALUED SEMANTICS

Ultimately, logical proportions, and in particular the homogeneous ones, could be used for practical applications where we have to deal with missing information or features whose satisfaction is a matter of degree. To cover such situations, extensions of the Boolean interpretation to multiple-valued logics (3-valued at least) is necessary. A formal way to cope with these situations is to extend the Boolean framework to a multiple-valued one by introducing truth values belonging to $[0, 1]$. We should carefully distinguish between three cases:

- when feature satisfaction is a matter of degree instead of being binary, i.e., the truth value of a given feature may be an *intermediate* value between 0 and 1.

- when a feature does not make sense for a given item, i.e., the feature is *non applicable* to it.

- when *information* about some features *is missing*, i.e., we have no clue about the truth value of some features for some items, and the corresponding truth value is not known, i.e., *unknown*.

At this stage, two questions arise:

1. in a given model, what are the valuations that correspond to a "perfect" proportion of a given type (i.e., having 1 as truth value)? For instance, does $T(a, a, a, a)$ postulate still have to be satisfied by A, R, P, or can we consider models where $A(u, u, u, u) = u$, u being a truth value distinct from 0 and 1?

2. are there valuations that could be regarded as "imperfect" proportions of a given type (i.e., with a truth value distinct from 0 and 1) and in that case, what is their truth value?

We investigate these issues in the following subsections keeping in mind an essential principle: whatever the way we define the truth values, the Boolean model should be the limit case of our models when restricted to Boolean valuations.

4.1 Semantics for gradual features

When the satisfaction of features may be a matter of degree, we have to consider that the truth values belong to a linearly ordered scale $\mathcal{L}$. The simplest case is when $\mathcal{L} = \{0, \alpha, 1\}$, with the ordering $0 < \alpha < 1$, which can be generalized into a finite chain $\mathcal{L} = \{\alpha_0 = 0, \alpha_1, \cdots, \alpha_n = 1\}$ of ordered grades $0 < \alpha_1 < \cdots < 1$, or to an infinite chain using the real interval $[0, 1]$. A proposal for extending A in such cases has been advocated in [Prade and Richard, 2010b]. It takes its source in the initial definition

$$A(a, b, c, d) = (a \wedge \bar{b} \equiv c \wedge \bar{d}) \wedge (\bar{a} \wedge b \equiv \bar{c} \wedge d),$$

where now

- i) the central $\wedge$ is taken as equal to min;

- ii) $s \equiv t$ is taken as $\min(s \rightarrow_L t, t \rightarrow_L s)$ where $\rightarrow_L$ is Łukasiewicz implication, defined by $s \rightarrow_L t = \min(1, 1 - s + t)$, for $\mathcal{L} = [0, 1]$ (in the discrete cases, we take $\alpha = 1/2$ and $\alpha_i = i/n$), and thus $s \equiv t = 1 - |s - t|$; note that $s \equiv t = (1 - s) \equiv (1 - t)$;

- iii) $s \wedge \bar{t} = \max(0, s - t) = 1 - (s \rightarrow_L t)$, i.e., $\wedge\,\bar{}$ is understood as expressing a bounded difference. Note that this choice preserves $A(a, b, c, d) = A(\bar{a}, \bar{b}, \bar{c}, \bar{d})$ for the involutive negation $\bar{x} = 1 - x$.

The resulting expression for $A(a,b,c,d)$ is given in Table 10. Then, we understand the truth value of $A(a,b,c,d)$ as the extent to which the truth values a,b,c,d makes an analogical proportion. For instance, in such a graded model, the truth value of $A(0.9,1,1,1) = 0.9$, which fits the intuition. It can be checked that the semantics of $A(a,b,c,d)$ thus defined in the graded case, reduces to the previous definition when restricted to the Boolean case.

It is interesting to study in what cases $A(a,b,c,d) = 1$ (and in what cases $A(a,b,c,d) = 0$). Then it is clear that $A(a,b,c,d) = 1$ when $a - b = c - d$. When $a,b,c,d \in \{0, \alpha = 1/2, 1\}$, it yields the 19 following patterns 1111; 0000; $\alpha\alpha\alpha\alpha$; 1010; 0101; $1\alpha1\alpha$; $\alpha1\alpha1$; $0\alpha0\alpha$; $\alpha0\alpha0$; 1100; 0011; $11\alpha\alpha$; $\alpha\alpha11$; $\alpha\alpha00$; $00\alpha\alpha$; $1\alpha\alpha0$; $0\alpha\alpha1$; $\alpha10\alpha$; $\alpha01\alpha$.

This means that $A(a,b,c,d) = 1$ when the change from a to b has the same direction and the same intensity as the change from c to d. However, the last 4 patterns show that there is no need to have $a = b$ and $a = c$ while these conditions hold for the 15 first patterns, which are all of the form $xyxy$, $xxyy$, or $xxxx$. In contrast, note that the last 4 patterns exhibit 3 distinct values.

Table 10. Graded definitions for A, R, P^*

$A(a,b,c,d) =$ $1- \mid (a-b)-(c-d) \mid$ if $a \geq b$ and $c \geq d$, or $a \leq b$ and $c \leq d$ $1 - max(\mid a-b\mid, \mid c-d\mid)$ if $a \leq b$ and $c \geq d$, or $a \geq b$ and $c \leq d$
$R(a,b,c,d) = A(a,b,d,c)$
$P^*(a,b,c,d) =$ $min(1 - \mid max(a,b) - max(c,d)\mid, 1 - \mid min(a,b) - min(c,d)\mid)$

$A(a,b,c,d) = 0$ when $a - b = 1$ and $c \leq d$, or $b - a = 1$ and $d \leq c$, or $a \leq b$ and $c - d = 1$, or $b \leq a$ and $d - c = 1$. It means the 22 following patterns in the 3-valued case: 1110; 1101; 1011; 0111; 0001; 0010; 0100; 1000; 1001; 0110; $10\alpha\alpha$; $01\alpha\alpha$; $\alpha\alpha10$; $\alpha\alpha01$; 100α; 011α; $10\alpha1$; $\alpha001$; $0\alpha10$; $1\alpha01$; $01\alpha0$; $\alpha110$. Thus, $A(a,b,c,d) = 0$ when one change inside the pairs (a,b) and (c,d) is maximal, while the other pair shows no change or a change in the opposite direction.

Using $\mathcal{L} = \{0, \alpha, 1\}$, $A(a,b,c,d) = \alpha$ for 81 - 19 - 22 = 40 distinct patterns.

In [Prade and Richard, 2010b], the graded extension of $R(a,b,c,d)$ is defined by permuting c and d in the definition of A, according to Proposition 4. But the extension of the paralogy is no longer obtained by permuting b

and d in the definition of A (as Proposition 4 would suggest). In fact, the paralogical proportion is defined directly from P^* (thus changing $\bar{a} \wedge \bar{b} \equiv \bar{c} \wedge \bar{d}$ into $a \vee b \equiv c \vee d$), and taking $\wedge = \min$, $\vee = \max$, and $s \equiv t = 1 - |s - t|$, we obtain the definition in Table 10. If we now exchange b and d (using Proposition 4 again) in this definition, we get the graded version of A^* (which is no longer equivalent to A), namely

$$A^*(a, b, c, d) = min(1 - |max(a, d) - max(b, c)|, 1 - |min(a, d) - min(b, c)|)$$

This is the direct counterpart of the definition without negation of the analogical proportion in the Boolean case. This alternative extension still preserves $A^*(a, b, c, d) = A^*(\bar{a}, \bar{b}, \bar{c}, \bar{d})$ for the involutive negation $\bar{x} = 1 - x$. It can be checked that $A^*(a, b, c, d) = 1$ only for the 15 patterns with at most two distinct values (for which $A(a, b, c, d) = 1$), while $A^*(a, b, c, d) = \alpha$ for the 4 other patterns for which $A(a, b, c, d) = 1$, namely for $1\alpha\alpha0$; $0\alpha\alpha1$; $\alpha10\alpha$; $\alpha01\alpha$. Besides, $A^*(a, b, c, d) = 0$ for only 18 among the 22 patterns that make $A(a, b, c, d) = 0$. The 4 patterns for which $A^*(a, b, c, d) = \alpha$ (instead of 0) are $10\alpha\alpha$; $01\alpha\alpha$; $\alpha\alpha10$; $\alpha\alpha01$.

Using $\mathcal{L} = \{0, \alpha, 1\}$, $A^*(a, b, c, d) = \alpha$ for 81 - 15 - 18 = 48 distinct patterns.

Thus, it appears that $A^*(a, b, c, d)$ does not acknowledge as perfect the analogical proportion patterns where the amount of change between a and b is the same as between c and d and has the same direction, but where this change applies in different areas of the truth scale. Still, $A^*(a, b, c, d)$ remains half-true in these cases, for $\mathcal{L} = \{0, \alpha, 1\}$. When $\mathcal{L} = [0, 1]$, it can be checked that $A^*(a, b, c, d) \geq 1/2$ when $a - b = c - d$; in particular, $\forall a, b, A^*(a, b, a, b) = 1$, which corresponds to the case where $a = c$ and $b = d$. In the same spirit, if $\mathcal{L} = \{0, \alpha, 1\}$ as well as for $\mathcal{L} = [0, 1]$, $A^*(a, b, c, d) = 0$ when a change inside the pairs (a,b) and (c,d) is maximal, while the other pair shows a change in the opposite direction starting from 0 or 1. However, $A^*(1, 0, c, c) = min(c, 1 - c)$ and A^* takes the same value for the 7 other permutations of $(1, 0, c, c)$ obtained by applying symmetry and/or central permutation.

As can be seen in Table 11, A^* and A also coincide on some patterns having intermediary truth values, but diverge on others. Generally speaking, A^* is smoother than A in the sense that more patterns have intermediary truth values with A^* than with A.

Both A and A^* continue to satisfy the *symmetry property* (as P, R, and P^*, R^* with $R^*(a, b, c, d) = A^*(a, b, d, c) = P^*(a, c, d, b)$). However, only A^*

Table 11. The two graded definitions of the analogical proportion in [0, 1]

A	A^*
$A(1,1,u,v) = 1 - \|u - v\|$	$A^*(1,1,u,v) = 1 - \|u - v\|$
$A(1,0,u,v) = u - v$ if $u \geq v$ $= 0$ if $u \leq v$	$A^*(1,0,u,v) = \min(u, 1-v)$
$A(0,1,u,v) = v - u$ if $u \leq v$ $= 0$ if $u \geq v$	$A^*(0,1,u,v) = min(v, 1-u)$
$A(0,0,u,v) = A(1,1,u,v)$	$A^*(0,0,u,v) = A^*(1,1,u,v)$

still enjoys the *means permutation* and the *extremes permutation* properties. *This is no longer the case with* A, as shown by the following counterexample.

$A(0.8, 0.6, 1, 0.3) = 1 - |(0.8 - 0.6) - (1 - 0.3)| = 1 - |0.2 - 0.7| = 0.5$ since $0.8 \geq 0.6$ and $1 \geq 0.3$, and $A(0.8, 1, 0.6, 0.3) = 1 - \max(|0.8 - 1|, |0.6 - 0.3|) = 1 - \max(0.2, 0.3) = 0.7$ since $0.8 \leq 1$ and $0.6 \geq 0.3$.

But, as already mentioned, *both* A *and* A^* continue to satisfy the *code independency* property with respect to $\overline{a} = 1 - a$. Some more Boolean properties that remain valid in the multiple-valued case are summarized in Table 12.

Table 12. Graded properties of A, A^*, R, P

Property name	Formal definition	Proportion
full identity	$T(a,a,a,a)$	A^*, A, R, P
reflexivity	$T(a,b,a,b)$	A^*, A, P
reverse reflexivity	$T(a,b,b,a)$	R,P
sameness	$T(a,a,b,b)$	A^*, A, R
symmetry	$T(a,b,c,d) \to T(c,d,a,b)$	A^*, A, R, P
permutation of means	$T(a,b,c,d) \to T(a,c,b,d)$	A^*
permutation of extremes	$T(a,b,c,d) \to T(d,b,c,a)$	A^*
all permutations	$\forall i, j, T(a,b,c,d) \to T(p_{i,j}(a,b,c,d))$	none
semi-mirroring	$T(a,b,\overline{a},\overline{b})$	R
exchange mirroring	$T(a,b,\overline{b},\overline{a})$	A
negation compatib.	$T(a,\overline{a},b,\overline{b})$	none

4.2 Dealing with non-applicable features

The abbreviation 'n/a' (for *non applicable*) is currently used in data tables when an attribute does not apply, when a feature does not make sense for a particular item. However, the extensive use of 'n/a' may be often ambiguous when it also appears in the same tables when information is *non available* for some attribute values of some items. Indeed one has to carefully distinguish the case where the feature does apply to the item, but it is not known if the feature is true or is false for the item, from the case where the feature is neither true nor false for the item since the feature does not apply to it. The case of unknown truth values is discussed in the next section, while we now address the problem of dealing with genuinely non applicable features.

The idea of introducing a third truth value for 'non applicable' (na for short in the following) in the context of analogy can be already found in the pioneering work of Sheldon Klein [1982; 1983], which we already mentioned in the equation solving subsection 3.4. However, his handling of na is based on $(na \equiv na) = na$, which suggests that the evaluation of an analogical proportion where na appears may receive the truth value na, which is more in the spirit of understanding na as 'not available', or 'unknown'.

Indeed, although a property may be 'true', 'false', or 'non applicable' for an item, it seems natural to expect that $A(a, b, c, d)$ can only be 'true' or 'false', since $1na1na$ looks intuitively satisfactory as an analogical proportion, while $1na00$ is not. More precisely, in the context of non applicable properties, we have only 3 valuation patterns that should make an analogical proportion true: $xxxx$, $xyxy$, and $xxyy$, where $x, y \in \{0, 1, na\}$. Any other option should make it false, since $\{0, 1, na\}$ play the same role. This leads to acknowledge as true the 15 following valuations:

- 1111; 0000; $nananana$ corresponding to $xxxx$;

- 1010; 0101; $1na1na$; $na1na1$; $0na0na$; $na0na0$ corresponding to $xyxy$ with $x \neq y$;

- 1100; 0011; $11nana$; $nana11$; $nana00$; $00nana$ corresponding to $xxyy$ with $x \neq y$.

All the remaining valuations lead to false.

In other words, we are in a situation somewhat similar to the one encountered in the previous section in the case of a unique intermediary truth-value α between true and false, meaning 'half-true' (or equivalently 'half-false'), when we refuse the four valuations $1\alpha\alpha0$, $0\alpha\alpha1$, $\alpha01\alpha$ and $\alpha10\alpha$ as being true, *except that* now no valuation leads to the third truth value. It is possible to find logical definitions of the analogical proportion having the expected behavior for the truth values $\{0, 1, na\}$. A solution to get the exact

truth table is:
- to order $\{0, 1, na\}$ as the chain $1 > na > 0$,
- to use Kleene conjunction and disjunction, see, e.g., [Ciucci and Dubois, 2012], respectively defined by the minimum and the maximum according to the above ordering,
- to use the strong Kleene equivalence $\equiv$, where $x \equiv y = 1$ if and only if $x = y$, and $x \equiv y = 0$ otherwise,
- to define analogical proportion with A^* instead of A, namely

$$A^*(a, b, c, d) = (a \wedge d \equiv b \wedge c) \wedge (a \vee d \equiv b \vee c).$$

A counterpart to $A(a, b, c, d) = (a \setminus b \equiv c \setminus d) \wedge (b \setminus a \equiv d \setminus c)$ where $\setminus$ here denotes the Boolean logical connective corresponding to set difference, can also be found. However, since we do not want to have $1nana0$ true, the difference between 1 and na and the difference between na and 0 should not be the same, neither the same as between 1 and 0, nor between 1 and 1 for sure. Thus we need 4 distinct values for the difference. This is impossible with 3 truth values! This contrasts with the Boolean case where there are only two possible difference values needed. The solution is then to use *2* connectives for differences:

$x \setminus_1 y = 1$ if $x = 1$ and $y = 0$; $x \setminus_1 y = na$ if $x = 1$ and $y = na$; $x \setminus_1 y = 0$ otherwise;

$x \setminus_2 y = 1$ if $x = 1$ and $y = 0$; $x \setminus_2 y = na$ if $x = na$ and $y = 0$; $x \setminus_2 y = 0$ otherwise.

Then the definition of $A(a, b, c, d)$ becomes

$$(a \setminus_1 b \equiv c \setminus_1 d) \wedge (b \setminus_2 a \equiv d \setminus_2 c) \wedge (a \setminus_2 b \equiv c \setminus_2 d) \wedge (b \setminus_1 a \equiv d \setminus_1 c)$$

where $x \equiv y = 1$ iff $x = y$; $x \equiv y = 0$ otherwise; and $\wedge$ is any conjunction connective that coincides with classical conjunction on $\{0, 1\}$. This definition yields 1 for the 15 expected patterns and is 0 otherwise for the 81 - 15 = 66 remaining patterns.

It is even possible to find an expression for $A(a, b, c, d)$ where $\setminus_1$ and $\setminus_2$ are expressed in terms of a conjunction (denoted $\wedge^*$) and two negations, i.e., where $x \setminus_1$ y is replaced by $x \wedge^* \bar{y}^1$ and $x \setminus_2 y$ is replaced by $x \wedge^* \bar{y}^2$. We obtain a definition for $A(a, b, c, d)$ under the form

$$(a \wedge^* \bar{b}^1 \equiv c \wedge^* \bar{d}^1) \wedge^* (b \wedge^* \bar{a}^2 \equiv d \wedge^* \bar{c}^2) \wedge^* (a \wedge^* \bar{b}^2 \equiv c \wedge^* \bar{d}^2) \wedge^* (b \wedge^* \bar{a}^1 \equiv d \wedge^* \bar{c}^1)$$

where the two negations are Post-like negations defined through a circular ordering of the three truth-values, where the negation of a value is the successor value in the ordering, namely $\overline{0}^1 = na; \overline{na}^1 = 1; \overline{1}^1 = 0$ and

$\overline{0}^2 = 1; \overline{na}^2 = 0; \overline{1}^2 = na$. This acknowledges the fact that in some sense these three truth-values play similar roles. The non-standard three-valued conjunction $\wedge^*$, which is defined by

$x \wedge^* y = 1$ if $x = 1$ and $y = 1$
$x \wedge^* y = u$ if $x = na$ and $y = na$
$x \wedge^* y = 0$ otherwise

also agrees with this view, while coinciding with classical conjunction in the binary case.

As in the previous section, we summarize in Table 13 the properties of the Boolean case that remain valid in this 3-valued model where na, standing for non applicable, is the third truth value.

Table 13. Properties of A, R, P with truth value na (as non applicable)

Property name	Formal definition	Proportion
full identity	$T(a, a, a, a)$	A,R,P
reflexivity	$T(a, b, a, b)$	A,P
reverse reflexivity	$T(a, b, b, a)$	R,P
sameness	$T(a, a, b, b)$	A,R
symmetry	$T(a, b, c, d) \to T(c, d, a, b)$	A,R,P
permutation of means	$T(a, b, c, d) \to T(a, c, b, d)$	A
permutation of extremes	$T(a, b, c, d) \to T(d, b, c, a)$	A
all permutations	$\forall i, j, T(a, b, c, d) \to T(p_{i,j}(a, b, c, d))$	none

4.3 Dealing with unknown features

In this section, we briefly consider a situation that is quite different from the ones studied in the two previous sections. We assume now that the features used for describing situations are all binary (i.e., they can be only true or false), but their truth value may be unknown.

Thus, the possible states of information regarding a Boolean variable x pertaining to a given feature may be represented by one of the 3 truth value subsets $\{0\}, \{1\}$ or $\{0, 1\}$, corresponding respectively to the case where the truth value of x is false, true or unknown. We denote this state of information by $\tilde{x}$, which is a subset of $\{0, 1\}$. The evaluation of a logical proportion $T(a, b, c, d)$ then amounts to compute the state of information denoted $\mathcal{T}(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d})$ about its truth value, knowing $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$. It is given by the standard set extension:

$$\mathcal{T}(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}) = \{v(T(a, b, c, d)) \mid v(a) \in \tilde{a}, v(b) \in \tilde{b}, v(c) \in \tilde{c}, v(d) \in \tilde{d}\}$$

where v denotes a Boolean valuation.

From now on, we focus on analogical proportion A only, but R, P and I could be handled in a similar manner. For instance, let us take the example $A(a, b, c, d)$ where $\tilde{a} = \{1\}, \tilde{b} = \{0\}, \tilde{c} = \tilde{d} = \{0, 1\}$. Applying the previous formula leads to

$$\mathcal{A}(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}) = \{0, 1\}$$

since the truth value of $A(a, b, c, d)$ may be 0 for the valuations 1001, 1000, 1011, and 1 for 1010.

Let us now consider the following expression $A(a, b, a, b)$ when $\tilde{a} = \tilde{b} = \{0, 1\}$. A similar computation leads to

$$\mathcal{A}(\tilde{a}, \tilde{b}, \tilde{a}, \tilde{b}) = \{1\}$$

since the truth value of $A(a, b, a, b)$ is 1 for any of the valuations 1010, 1111, 0101, or 0000. Similarly, the truth value of $A(a, a, a, a)$ is 1, even when $\tilde{a} = \{0, 1\}$.

But, the set of possible truth values for $A(a, b, c, d)$ is $\{0, 1\}$ when $\tilde{a} = \{0, 1\}, \tilde{b} = \{0, 1\}, \tilde{c} = \{0, 1\}, \tilde{d} = \{0, 1\}$. It should be clear that this does not mean that the Boolean variables a, b, c, d are equal; we just have the same state of information for all of them. This expresses that the full identity property does not hold any longer at the information level for analogical proportion. And this illustrates the fact that the logic of uncertainty is no longer truth functional, since the state of information about the truth value of $A(a, b, c, d)$ does not only depend on the state of information about the truth values of a, b, c, and d, but is also constrained by the existence of possible logical dependencies between these variables.

Nevertheless, some key properties of homogeneous proportions remain valid at the information level such as symmetry, or central and extreme permutations. Indeed it can be checked that, for instance, for symmetry:

$$\mathcal{A}(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}) = \mathcal{A}(\tilde{c}, \tilde{d}, \tilde{a}, \tilde{b})$$

Using the set extension evaluation of logical proportions in presence of incomplete information, we can compute the set of possible truth values of the analogical proportion for the different 4-tuples of states of information. We now denote by u the state $\{0, 1\}$, and respectively by 0 and 1, the states of information $\{0\}$ and $\{1\}$. A 4-tuple of states of information will be called *information pattern*, or pattern for short, and denoted by a 4-tuple of elements of $\{0, 1, u\}$ without blank space. For instance, $01u1$ is such a pattern and should be understood as the 4-tuple of states of information $(\{0\}, \{1\}, \{0, 1\}, \{1\})$.

Then, the 6 patterns $0000, 1111, 0011, 1100, 1010, 0101$ that makes A true in the Boolean case, and where u does not appear, are the only ones that are still true with the above view (for which we get the singleton $\{1\}$ as information state for $A(a,b,c,d)$). As soon as at least one state of information is u in the pattern, the state of information for $A(a,b,c,d)$ is u or 0. Indeed, for instance, $01u0$ leads to 0 since whatever the truth value of the 3rd variable, the analogical proportion does not hold. Thus, despite the lack of knowledge regarding the 3rd variable, we know the exact truth value of the proportion in this case, namely it is false. It appears that there are 18 patterns that lead to 0. They are the 10 patterns of the Boolean case and the 8 following ones: $01u0$, $0u10$, $u001$, $100u$, $10u1$, $1u01$, $u110$, $011u$. Thus, in the $81 - 6 - 18 = 57$ remaining cases, the state of information for $A(a,b,c,d)$ is u.

It can be checked that these results can be retrieved both with the initial definition of A or with A^* where complete ignorance u is handled with $^-$, $\wedge$, $\vee$ as the strong Kleene connectives (see [Ciucci and Dubois, 2012]) and $\equiv$ as Bochvar connective, where u is an absorbing element. The corresponding truth tables are recalled in Table 14. This provides a way to extend the

Table 14. Truth tables for u as lack of knowledge

	$^-$	$\wedge$	0	1	u	$\vee$	0	1	u	$\equiv$	0	1	u
0	1	0	0	0	0	0	0	1	u	0	1	0	u
1	0	1	0	1	u	1	1	1	1	1	0	1	u
u	u	u	0	u	u	u	u	1	u	u	u	u	u

definition of the analogical proportion in case of lack of knowledge when no dependencies between the variables exist. As in the Boolean case, the definitions A (resp. R, P, I) and A^* (resp. R^*, P^*, I^*) are equivalent.

Nevertheless, this truth-functional calculus provides only a description of the evaluation of the patterns at the information level. Namely, it enables us to retrieve the tri-partition of the patterns in respectively 6, 18 and 57 patterns leading respectively to 1, 0 and u, but it does not account for the full calculus of the extended definition of logical proportions in presence of incomplete information, when dependencies take place between variables, for instance it can be checked that $A(a,b,a,b)$ and $A^*(a,b,a,b)$ when a and b are unknown does not yield 1 as expected, but u (this is just due to the fact that constraints $a = c$ and $b = d$ are ignored).

5 HETEROGENEOUS PROPORTIONS

As highlighted in the introduction, there are 4 other proportions that satisfy code independency, and as such stand out of the 120 logical proportions, namely the heterogeneous proportions, whose truth tables are given in Table 15.

Table 15. H_a, H_b, H_c, H_d - Boolean truth tables

H_a				**H_b**				**H_c**				**H_d**			
1	1	1	0	1	1	1	0	1	1	1	0	1	1	0	1
0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0
1	1	0	1	1	1	0	1	1	0	1	1	1	0	1	1
0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0
1	0	1	1	0	1	1	1	0	1	1	1	0	1	1	1
0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0

It is stunning to note that these truth tables exactly involve the 8 missing tuples of the homogeneous tables, i.e., those ones having an odd number of 0 and 1. It should not come as a surprise that they satisfy the same association properties as the homogeneous ones: for instance, any combination of 2 or 3 heterogeneous proportions is satisfiable, but the conjunction $H_a(a, b, c, d) \wedge H_b(a, b, c, d) \wedge H_c(a, b, c, d) \wedge H_d(a, b, c, d)$ is not satisfiable. This fact contributes to make the heterogeneous proportions the perfect dual of the homogeneous ones.

5.1 Properties

The formal definitions given in Table 2 lead to immediate Boolean equivalences between heterogeneous and homogeneous proportions that we summarize in Table 16.

Obviously, the heterogeneous proportions are strongly linked together: for instance, using the symmetry of I,

$$H_a(a, b, c, d) \equiv I(\overline{a}, b, c, d) \equiv I(c, d, \overline{a}, b) \equiv H_c(c, d, a, b).$$

We may consider two different ways for generating these proportions:

- A *semantic viewpoint*: The *full identity* postulate $T(a, a, a, a)$ asserts that proportion T holds between identical values. Negating one variable position only generates an intruder, as in $T(\overline{a}, a, a, a)$, $T(a, \overline{a}, a, a)$, $T(a, a, \overline{a}, a)$ and $T(a, a, a, \overline{a})$, and leads to new postulates respectively denoted T_a, T_b, T_c and T_d. We call the negated position the *intruder position*: for instance, T_a

Table 16. Equivalences between heterogeneous and homogeneous proportions

$\mathbf{H_a}$	$\mathbf{H_b}$
$H_a(a,b,c,d) \equiv I(\overline{a},b,c,d)$	$H_b(a,b,c,d) \equiv I(a,\overline{b},c,d)$
$H_a(a,b,c,d) \equiv P(\overline{a},b,\overline{c},\overline{d})$	$H_b(a,b,c,d) \equiv P(a,\overline{b},\overline{c},\overline{d})$
$H_a(a,b,c,d) \equiv P(a,\overline{b},c,d)$	$H_b(a,b,c,d) \equiv P(\overline{a},b,c,d)$
$\mathbf{H_c}$	$\mathbf{H_d}$
$H_c(a,b,c,d) \equiv I(a,b,\overline{c},d)$	$H_d(a,b,c,d) \equiv I(a,b,c,\overline{d})$
$H_c(a,b,c,d) \equiv P(\overline{a},\overline{b},\overline{c},d)$	$H_d(a,b,c,d) \equiv P(\overline{a},\overline{b},c,\overline{d})$
$H_c(a,b,c,d) \equiv P(a,b,c,\overline{d})$	$H_d(a,b,c,d) \equiv P(a,b,\overline{c},d)$

expresses the fact that the first position is an intruder. For a proportion, to satisfy the property T_a means that *the first variable may be an intruder*. Since each postulate T_a, T_b, T_c and T_d is validated by 2 distinct valuations, it is clear that 3 of them are enough to define a logical proportion having exactly 6 valid tuples. There is no proportion satisfying all these postulates since it leads to 8 valid tuples, which excludes any logical proportion. It can be easily checked that H_a satisfies T_b, T_c, T_d and does not satisfy T_a: then H_a is uniquely characterized by the conjunction of properties $T_b \wedge T_c \wedge T_d$. We can interpret $H_a(a,b,c,d)$ as the following assertion: *the first position is not an intruder and there is an intruder among the remaining positions*. As a consequence, $H_a(a,b,c,d)$ does not hold when there is no intruder (i.e., when there is an even number of 0), or when a is the intruder. The same reasoning applies to H_b, H_c, H_d.

- A *syntactic viewpoint*: Here we start from the definition of the inverse paralogy I: $(a \wedge b \equiv \overline{c} \wedge \overline{d}) \wedge (\overline{a} \wedge \overline{b} \equiv c \wedge d)$. To get the definition of an heterogeneous proportion satisfying postulates where the intruder is in position 4, 2 or 1 for instance, we add a negation on the 3rd variable in both equivalences defining I. Here we get H_c as:

$$(a \wedge b \equiv c \wedge \overline{d}) \wedge (\overline{a} \wedge \overline{b} \equiv \overline{c} \wedge d)$$

This process, allowing us to generate the 4 heterogeneous proportions, shows that, in some sense, they are "atomic perturbations" of I: for this reason and since they are heterogeneous proportions, they have been respectively denoted H_a, H_b, H_c and H_d where the subscript corresponds to:

- the postulate *which is not satisfied* by the corresponding proportion or, equivalently,
- the negated variable in the equivalence with I.

For instance $H_a(a,b,c,d) \equiv I(\overline{a},b,c,d)$, H_a satisfies T_b, T_c, T_d and does not satisfy T_a. This leads to another way to interpret $H_a(a,b,c,d)$. Since $H_a(a,b,c,d) \equiv I(\overline{a},b,c,d)$, when $H_a(a,b,c,d) = 1$, a is not the intruder, i.e., $\overline{a}$ is the value of the intruder. The different possible cases are as follows:

- $\overline{a}bcd$= 1100 or 0011 and the intruder is b.
- or $\overline{a}bcd$ = 0101, 0110, 1010 or 1001 and the intruder is c or d.

In other words, there is an intruder in (a,b,c,d), which is not a, iff the properties common to $\overline{a}$ and b (positively or negatively) are not those common to c and d, and conversely.

As in the case of homogeneous proportions, the semantic properties of heterogeneous proportions are easily derived from their truth tables, which we summarize in Table 17. It is clear on their truth tables, that none of the

Table 17. Properties of heterogeneous proportions

Property name	Formal definition	Proportion
full identity	$T(a,a,a,a)$	none
reflexivity	$T(a,b,a,b)$	none
reverse reflexivity	$T(a,b,b,a)$	none
sameness	$T(a,a,b,b)$	none
symmetry	$T(a,b,c,d) \to T(c,d,a,b)$	none
means permut.	$T(a,b,c,d) \to T(a,c,b,d)$	H_a, H_d
extremes permut.	$T(a,b,c,d) \to T(d,b,c,a)$	H_c, H_b
all permutations	$\forall i,j, T(a,b,c,d) \to T(p_{i,j}(a,b,c,d))$	none
transitivity	$T(a,b,c,d) \wedge T(c,d,e,f) \to T(a,b,e,f)$	none
Ta	$T(\overline{a},a,a,a)$	H_b, H_c, H_d
Tb	$T(a,\overline{a},a,a)$	H_a, H_c, H_d
Tc	$T(a,a,\overline{a},a)$	H_a, H_b, H_d
Td	$T(a,a,a,\overline{a})$	H_a, H_b, H_c

heterogeneous proportions satisfy neither symmetry nor transitivity. From a practical viewpoint, these proportions are closely related with the idea of spotting the odd one out (the intruder), or if we prefer of picking the one that doesn't fit among 4 items. This will be further discussed in Section 6, but we first consider the extension of heterogeneous proportions to the case of graded properties with intermediate truth values.

5.2 multiple-valued semantics

We extend here what has been done for homogeneous proportions and their multiple-valued semantics in Section 4.1. Roughly speaking, in the case of H_a, the graded truth value of $H_a(a, b, c, d)$ estimates how *far* we are from having a as an intruder.

Obviously the same questions as for homogeneous proportions arise but with a different interpretation:

1) what are the valuations that correspond to a "perfect" proportion of a given type (i.e., having 1 as truth degree)? For instance, we want the truth value of $H_a(0, u, 0, u)$ to be equal to 1 (as well as the truth value of $H_c(0, u, 0, u)$) because in that context, it is true that $a = 0$ (resp. c) cannot be the intruder, whatever the value of u.

2) are there valuations that could be regarded as *approximate* proportions of a given type (with an intermediate truth degree) and in that case, what is their truth value? For instance, in the valuation $(0.7, 1, 1, 0.9)$, it is likely that a is the intruder just because the other candidate, d, has a value very close to 1, and the closer d is to 1, the more likely a is the intruder: then the truth value of $H_a(0.7, 1, 1, 0.9)$ should be small and related to $1 - d = 0.1$ (since H_a excludes a as intruder).

The most rigorous way to proceed is to start from the definition of multiple-valued paralogy given in Section 4.1. This definition is based on P^*: it leads, for a three valued scale, to 15 valuations fully true, and 18 fully false. The 48 remaining patterns get intermediate truth value given by the following general formula

$$P^*(a, b, c, d) = min(1-|max(a, b)-max(c, d)|, 1-|min(a, b)-min(c, d)|)$$

which, thanks to the symmetry of P^* and stability w.r.t. the permutation of its two first variables, has the following behavior

general case	case $u = v$
$P^*(1, 1, u, v) = \min(u, v)$	$P^*(1, 1, u, u) = u$
$P^*(1, 0, u, v) = \min(\max(u, v), 1-\min(u, v))$	$P^*(1, 0, u, u) = \min(u, 1-u)$
$P^*(0, 0, u, v) = 1 - \max(u, v)$	$P^*(0, 0, u, u) = 1 - u$

Starting from the equivalences given in Table 16, we get the multi-valued definition for H_a (and similar definitions for H_b, H_c, H_d), still leading to 15 true valuations, 18 false valuations and 48 with intermediate values in case of a three valued scale:

$$H_a(a, b, c, d) = \min(1-|\max(a,1-b)-\max(c,d)|, 1-|\min(a,1-b)-\min(c,d)|)$$

Let us note that $H_a(0,0,u,v) = H_a(1,1,u,v)$ due to the equality $H_a(0,0,u,v) =$

$P(0,1,u,v) = P(1,0,u,v)$. We have

general case	case $u = v$
$H_a(1,1,u,v) = \min(\max(u,v), 1-\min(u,v))$	$H_a(1,1,u,u) = \min(u,1-u)$
$H_a(1,0,u,v) = \min(u,v)$	$H_a(1,0,u,u) = u$
$H_a(0,1,u,v) = 1 - \max(u,v)$	$H_a(1,0,u,u) = 1 - u$

Let us analyze two examples to highlight the fact that the above definition really fits with the intuition.

- Considering the valuation $100u$, its truth value is:

  - u for P: if u is close to 1, we are close to the fully true paralogical proportion and the truth value is high. In the opposite case, u is close to 0 and we are close to a fully false paralogical proportion 1000.

  - 1-u for H_b, H_c, H_d: if u is close to 1, we are close to the valuation 1001 which is definitely not a valid valuation for H_b, H_c, H_d: so 1-u is a low truth value. But if u is close to 0, we are close to the valuation 1000 which is valid for H_b, H_c, H_d and 1-u is a high truth value.

  - finally 0 for H_a: whatever the value of u, 100u means "an intruder is in first position", when the semantics of H_a is just the opposite.

- Back to the graded valuation valuation 0.7 1 1 0.9 considered above

  - regarding P, the truth value as given by the formula is 0.8, i.e., the valuation is close to be a true paralogy.

  - regarding the heterogeneous proportions, we understand that we have 2 candidate intruders namely $a = 0.7$ and $d = 0.9$. But they are not equivalent in terms of intrusion and it is more likely to be a than d. This is consistent with the fact that the truth value of $H_a(0.7, 1, 1, 0.9)$ is 0.1 (very low), but the truth value of $H_d(0.7, 1, 1, 0.9)$ is 0.3 (a bit higher).

  - in fact, 0.7 1 1 0.9 does not give a genuine impression that there is an intruder, which is in agreement with the fact that $H_b(0.7, 1, 1, 0.9) = H_c(0.7, 1, 1, 0.9) = 0.4$.

6 APPLICATIONS

In this section, we provide an overview of the use of logical proportions for various reasoning purposes. Since we have distinguished two remarkable groups of proportions, differing both from a syntactic and a semantic viewpoint, it is not surprising that they can be used for two different styles of

Figure 1. IQ test: Graphical analogy

applications. On the one hand, the homogeneous proportions allow us to build up a missing item in a given sequence. On the other hand, the heterogeneous proportions are suitable for a dual task which is to pick up the odd one out in a set. Let us start by discussing the use of the homogeneous logical proportions, which is the most developed.

6.1 Using homogeneous proportion for finding missing values

From a general viewpoint, a homogeneous proportion between 4 items a, b, c, d expresses that the elements of the pair (a, b) are similar (or dissimilar) in a way that can be related to the way the elements of the pair (c, d) are similar (or dissimilar). The equation-solving process described above enables us to compute d from the knowledge of a, b, c, when possible. Obviously, in practical cases, the items to be considered cannot be simply described by a single Boolean (or multiple-valued) variable, and a straightforward extension, allowing to cope with more sophisticated representations, can be given for Boolean vectors in $\mathbb{B}^n$, as follows (where T denotes any logical proportion):

$$T(\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}, \overrightarrow{d}) \text{ iff } \forall i \in [1, n],\ T(a_i, b_i, c_i, d_i).$$

The solving process of the equation $T(\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{c}, \overrightarrow{x})$ is still effective: instead of getting one Boolean value, we get a Boolean vector, by solving equations componentwise, computing d_i from a_i, b_i, and c_i (provided that the solution exists). This can be illustrated on a sequence of 3 pictures to be completed (see Figure 1, as it is often the case in IQ tests . Indeed, a noticeable part of the IQ tests are based on providing incomplete analogical proportions (see, e.g., [French, 2002]). Usually, the 3 first items A, B, C are given and the 4th item X has to be chosen among several plausible candidates. In this case, the homogeneous logical proportion method applies straightforwardly. The items A, B, C in the example of Figure 1 can be described respectively by vectors $(1, 0, 1, 0, 1)$, $(1, 0, 0, 1, 1)$, $(0, 1, 1, 0, 1)$, where the vector components refer respectively to the presence (or not) of a square, of a triangle, of a star, of a circle, and of a black point. Assuming that an analogical proportion

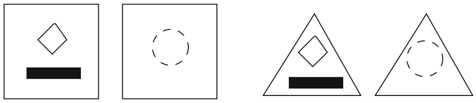


Figure 2. Analogy with a graded feature

should hold, by solving componentwise the analogical proportion equations expressing that $A(a_i, b_i, c_i, x_i)$ holds true for $i = 1, 5$, we easily get $X = (0, 1, 0, 1, 1)$, which corresponds to the result exhibited in Figure 1. Note that X is directly computed with this method, rather than chosen among a set of more or less "distant" potential solutions that would be given. In case the analogical equation has no solution for some component, on may try if another homogeneous proportion would fit for all the features. It would not be difficult to build examples of sequences of 4 pictures, where the display of squares, triangles, stars, circles and black dots is different from Figure 1, and where the fourth picture would be obtained via one of the three other homogeneous proportions R, P, or I, rather than via A as in Figure 1.

Moreover, Figure 2 illustrates the idea of having graded features, where here the presence of a circle is a matter of degree (the more densely dotted the circle, the higher the degree α of presence of a circle (in Figure 2 the analogical proportion $A(0, \alpha, 0, \alpha)$ clearly holds for the 'circle' variable).

In the above example, the problem is handled at a rather high conceptual level that requires that triangles, circles and so on be identified in the pictures. However, it has been pointed out [Prade and Richard, 2011] that the analogical proportion-based technique can still be applied at the pixel level. Then a black and white picture is represented by the Boolean vector made of its *bitmap* description that acknowledges (or not) the black color of each pixel. This supposes that all the geometric shapes (squares, triangles, stars, circles) use exactly the same pixels in all cases. Then, the proportion-based procedure automatically builds the associated geometric figure (when it exists), without introducing any knowledge about triangle, circle, etc.

Lastly, let us mention that it may be convenient to have extensions of the proportions allowing for the explicit handling of functional symbols, as in, e.g., the analogical proportion $A(x, f(x), y, f(y))$, for handling more sophisticated sequences of pictures (where for instance, elements are reversed from one picture to another), or analogical proportions quizzes like "*abc* is to *abd* as ijk is to ?" (where we have to encode that d is the successor of c); see [Correa *et al.*, 2012].

6.2 Classification and matrix abduction

We now consider variants of the process described in the previous subsection, when it is first checked that an homogeneous proportion holds on a series of n features between 4 items, and on this basis, one extrapolates that the same logical proportion still holds for a $(n+1)$th feature of interest, which is known only for the first 3 items. Solving the logical proportion equation corresponding to this latter feature then enables us to compute a plausible value of this feature for the 4th item. Classification problems are an important instance of this situation where the $(n+1)$th feature refers to the class of the item while the n other features pertain to its description. See [Miclet *et al.*, 2008] for the case of binary features, where very good results are reported on classification benchmarks. The graded version A has been used for handling numerical features in classification problems (also with promising experimental results [Prade *et al.*, 2012]), while A^* has not been experienced yet. It is still unclear if A^* may be more suitable for classification purposes.

The problem of completing a matrix where some values are missing is quite close to the classification problem, and thus different methods may be thought of in order to deal with this issue. Whatever the technique, the main question is to know if the extra knowledge that we may have about the problem and the available data carry sufficient information for an accurate reconstruction of the missing cells. This is not always the case. We focus here on a particular case, called "matrix abduction problem", using [Abraham *et al.*, 2009]'s terminology. It consists in guessing plausible values for cells having empty information in a matrix where each line corresponds to a situation described according to different binary features (each column corresponds to a particular feature).

Let us consider the screen example used by [Abraham *et al.*, 2009], where computer screens are described by 6 characteristic features: P is for price over £450, C for self collection, I for screen bigger than 24 inch, R for reaction time below 4ms, D for dot size less than 0.275, and S for stereophonic; 1 means "yes" and 0 means "no". We have 3 screens (screen 1, screen 2 and screen 4) whose characteristics are known and screen 3 where the truth value of the attribute S is missing (see Table 18).
To tackle such a common sense problem, a general idea (which may be also found in classification) is that replacing an unknown value by either 1 or 0 should result in the least possible *perturbation* of the matrix. This idea may be implemented in diverse ways. In [Abraham *et al.*, 2009] the idea is to choose the value that least perturbs the pre-existing partial order-

Table 18. The screen example

	P	C	I	R	D	S
$screen1$	0	1	0	1	0	1
$screen2$	0	0	1	1	0	1
$screen3$	0	0	0	0	1	?
$screen4$	1	1	0	0	1	1

ing between the column vectors of the matrix. In [Schockaert and Prade, 2011], the idea is rather to respect betweenness and parallelism relations that hold in conceptual spaces. We suggest here to enforce an homogeneous proportion T that already holds for completely known features.

Assume we have a Boolean vector incompletely describing a situation with respect to a set of $n+1$ considered features, say $v = (v_1, ..., v_n, x_{n+1})$, where for simplicity we assume that only x_{n+1} is unknown. For trying to make a plausible guess of the value of x_{n+1}, we have a collection (which may be rather small) of completely informed examples $e^i = (e^i_1, ..., e^i_n, e^i_{n+1})$ for $i = 1, n$. Then one may have at least three strategies:

i) comparing v to each e^i separately, and using a k-nearest neighbors approach, extending the idea that $T(e, e, e, v)$ should hold true and has $v = e$ as solution.

ii) looking for pairs e^i, e^j such that $T(e^i_h, v_h, v_h, e^j_h)$ makes a continuous homogeneous proportion T for a maximal number of features h, implementing the idea of having v_h between e^i_h and e^j_h ; observe however, that in the Boolean case, this would force to have the trivial situations $T(1, 1, 1, 1)$ or $T(0, 0, 0, 0)$ on a maximal number of features, and tolerate some "approximate" patterns $T(1, 1, 1, 0)$, $T(0, 1, 1, 1)$, $T(0, 0, 0, 1)$, or $T(1, 0, 0, 0)$, while rejecting patterns $T(0, 1, 1, 0)$ and $T(1, 0, 0, 1)$.

iii) looking for triples e^i, e^j, e^k such that $T(e^i_h, e^j_h, e^k_h, v_h)$ makes an homogeneous proportion T for a maximal number of features h.

In cases ii) or iii), the principle amounts to say that if an homogeneous proportion holds for a number of features as great as possible among features h such that $1 \leq h \leq n$, it should still hold for feature $n + 1$, which provides an equation for finding x_{n+1} if solvable. If there are several triples that are equally good in terms of numbers of features for which the proportion holds, but lead to different predictions, one may then consider that there is no acceptable plausible value for x_{n+1}.

The application of the first strategy on the above example yields 1 considering that screen 3 is already identical to screen 4 on 3 features. Using the second strategy, we observe that screen 3 is only in "between" screen 2 and screen 4 in the sense described above, leading again to 1 as a solution.

Using the third strategy that should involve 4 distinct items, we can observe that the analogical proportion $A(\mathit{screen}\ 1, \mathit{screen}\ 2, \mathit{screen}\ 4, \mathit{screen}\ 3)$ holds componentwise for features C, R, and D (while it fails with proportions P and I). Again we get 1 as a solution for ensuring an analogical proportion (namely $A(1,1,1,1)$) on S. Observe also that whatever the order in which the screens are considered, an homogeneous proportion holds for features C, R, D, and S. Considering other triples (if available) may lead to other equations having 0 as a solution. A prediction based on the triple making an homogeneous proportion with the incompletely described item on a maximal number of features, should be preferred. In case of ties on this maximal number of features between concurrent triples leading to opposite predictions, no prediction can be given. It is worth noticing that in [Abraham *et al.*, 2009], the use of 0 and 1 in the Boolean coding in their matrices is not just a matter of convention and we cannot exchange the 2 values since it will change the ordering. This is not the case with our approach since A, R, P, I satisfy code independency. The screen example is clearly a toy example but, in [Abraham *et al.*, 2009], similar examples are discussed which could also be handled using homogeneous proportions.

6.3 Analogical proportions in Raven's tests

Among the picture-based IQ tests (the use of pictures avoids the bias of a cultural background), the so-called Raven's Progressive Matrices [Raven, 2000] are considered as a reference for estimating the reasoning component of "the general intelligence". Recently a computational model for solving Raven's Progressive Matrices has been investigated in [Lovett *et al.*, 2010]. This model combines qualitative spatial representations with analogical comparison via structure-mapping [Gentner, 1983]. In the following, we suggest with an example that the Boolean proportion approach can be also used for solving such a test (see [Prade and Richard, 2011; Correa *et al.*, 2012] for other examples).

Each Raven test is constituted with a 3x3 matrix $pic[i, j]$ of pictures where the last picture $pic[3, 3]$ is missing and has to be chosen among a panel of 8 candidate pictures. An example is given in Figure 3 and its solution in

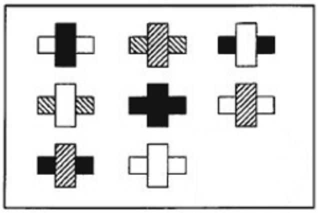


Figure 3. An example of Raven matrix

Figure 4. Raven matrix: the solution

Figure 4. We assume that the Raven matrices can be understood in the following way, with respect to rows and columns:

$$\forall i \in [1,2], \exists f \text{ such that } pic[i,3] = f(pic[i,1], pic[i,2])$$

$$\forall j \in [1,2], \exists g \text{ such that } pic[3,j] = g(pic[1,j], pic[2,j])$$

The two complete rows (resp. columns) are supposed to help to discover f (resp. g) and to predict the missing picture $pic([3,3])$ as $f(pic[3,1], pic[3,2])$ (resp. $g(pic[1,3], pic[2,3])$).

Obviously, these tests do not fit the standard equation solving scheme, but they follow an extended one telling us that $A((a,b), f(a,b), (c,d), f(c,d))$ holds for lines and $A((a,b), g(a,b), (c,d), g(c,d))$ for columns, i.e.

$$A((pic[1,1], pic[1,2]), pic[1,3], (pic[2,1], pic[2,2]), pic[2,3])$$

$$A((pic[1,1], pic[2,1]), pic[3,1], (pic[1,2], pic[2,2]), pic[3,2])$$

Thus, in that case, we have to consider a pair of cells $(pic[i,1], pic[i,2])$ as the first element of an analogical proportion, and then the pair $((pic[i,1], pic[i,2]),$ $pic[i,3])$ provides the 2 first element a and b of the analogical proportion we are considering. In terms of coding, in the example of Figure 3, we may consider the pictures as represented by a pair (or vector) (hr, vr) with one horizontal rectangle hr and a vertical one vr, each of these rectangles having one color among $Black, White, Grey$, we have then the following obvious encoding of the matrix in Table 19.

Table 19. Raven matrix: a coding

	1	2	3
1	WB	GG	BW
2	GW	BB	WG
3	BG	WW	?i?ii

It leads to the following analogical patterns (using the traditional notation for analogical proportion $a : b :: c : d$ instead of $A(a, b, c, d)$):

(WB,GG) : BW :: (GW, BB) : WG (1st and 2nd rows)
(WB,GG) : BW :: (BG,WW) : ?i?ii (1st and 3rd rows)
where BW = f(WB,GG) and WG = f(GW,BB).

(WB,GW) : BG :: (GG, BB) : WW (1st and 2nd columns)
(WB,GW) : BG :: (BW,WG) : ?i?ii (1st and 3rd columns)
where BG = f(WB,GW) and WW = g(GG,BB),

or if we prefer, since analogical proportions holds componentwise, we have the following valid proportions

- for the horizontal bars:
(W,G) : B :: (G, B) : W (horizontal analysis)
(W,G) : B :: (B,W) : ?i (horizontal analysis)
(W,G) : B :: (G, B) : W (vertical analysis)
(W,G) : B :: (B,W) : ?i (vertical analysis)
- for the vertical bars:
(B,G) : W :: (W, B) : G (horizontal analysis)
(B,G) : W :: (G,W) : ?ii (horizontal analysis)
(B,W) : G :: (G, B) : W (vertical analysis)
(B,W) : G :: (W,G) : ?ii (vertical analysis)

One can observe that the item (B, W) appears only in the analogical proportions with question marks for horizontal bars, while the items (G, W) and (W, G) appear only in the analogical proportions with question marks for vertical bars. Analogical proportions coming from both horizontal or vertical analysis are insufficient for concluding here. However, we can consider the Raven matrix provides a set of analogical associations without any distinction between those ones coming from the horizontal bars and those ones coming from vertical bars. In other words, we now relax the componentwise reading by considering that what applies to horizontal bars, may apply to vertical bars, and vice-versa. With this viewpoint, it appears that the pair (B, W) and the pair (W, G) are respectively associated to G (vertical association for vertical bar) and B (horizontal association for horizontal bar), which encodes the expected solution GB (as pictured in Figure 4). Note also that (G, W) cannot help predicting ?ii.

6.4 *Using heterogeneous proportions "to pick up the one which does not fit"*

As it is the case for homogeneous proportions, heterogeneous proportions can also be related to the solving of some type of quiz problem. As we have seen, the truth tables of the heterogeneous proportions highlight a Boolean value (0 or 1) which is different from the 3 remaining ones. It is then natural to think in terms of exception or intruder in a sequence of 4 items: the heterogeneous proportions play a dual role with regard to homogeneous proportions. Given a sequence of objects, they allow to distinguish the object which does not follow the "logic" of the sequence. As a consequence, heterogeneous proportions are suitable for the "Finding the odd one out" problem where a complete sequence of items being given, we have to find the item that does not fit with the other ones and which is, in some sense, an intruder or an anomaly. On this basis, a complete battery of IQ tests has been recently proposed in [Hampshire, 2010]. Solving 'Find the odd one out' tests (which are visual tests) has been recently tackled in [McGreggor and Goel, 2011] by using analogical pairing between fractal representation of the pictures. It is worth noticing that the approach of these authors takes its root in the idea of analogical proportion. However, this method relies on the use of similarity/dissimilarity measures rather than referring to a formal logical view of analogical proportion. In the following, we show that an opposite type (in some sense) of proportions, namely heterogeneous proportions, provides a convenient way to code and to tackle this problem.

Let us first consider the case of 4 items: obviously, if these items are completely different in many respects, there is no notion of intruder. The intruder comes as soon as there is a kind of unique dissimilarity among an obvious set of similarities or identities. Let us start with a simple case where each item a can be represented as a Boolean vector $a_1, \ldots, a_n$ where n is the number of attributes and $a_i \in \{0, 1\}$. Let us consider the simple example $(lorry, bus, bicycle, car)$ (where the obvious intruder is *bicycle*) shown in Figure 5 where $n = 5$ with a straightforward coding. When considering the item componentwise, we see that:

	canMove	hasEngine	onRoad	has4Wheel	canFly
lorry	1	1	1	1	0
bus	1	1	1	1	0
bicycle	1	0	1	0	0
car	1	1	1	1	0

Figure 5. A simple quiz and its Boolean coding

- for $i = 1, 3, 5$, $H_a(a_i, b_i, c_i, d_i) = H_b(a_i, b_i, c_i, d_i) = H_c(a_i, b_i, c_i, d_i) = H_d(a_i, b_i, c_i, d_i) = 0$.
- for $i = 2, 4$, $H_a(a_i, b_i, c_i, d_i) = H_b(a_i, b_i, c_i, d_i) = H_d(a_i, b_i, c_i, d_i) = 1$.
- for $i = 2, 4$, $H_c(a_i, b_i, c_i, d_i) = 0$

The indexes $1, 3$ and 5 are not useful to pick up the intruder because all the proportions have the same truth value. This is not the case for the indexes 2 and 4: H_a for instance, being equal to 1, insures that there is an intruder (which is not the first element). The intruder is then given by the proportion having the value 0: for instance, $H_c(a_i, b_i, c_i, d_i) = 0$ means that the fact that c is not an intruder is false, which exactly means that c is the intruder for component j. In our example, H_c is 0 on both components 2 and 4: this exactly leads to consider the third element *bicycle*, intruder for the components 2 and 4, as the global intruder. It may be the case that, we do not get the same intruder depending on the component: in that case, a majority vote may be applied and we choose as intruder the one which is intruder for the maximum number of components.

Thanks to the multiple-valued extension, this method can be generalized to the non Boolean case where each item a is represented as a real vector $a_1, \ldots, a_n$ and $a_i \in [0, 1]$. Then, the truth values of H_a, H_b, H_c and H_d on some features may be close to 0, which means that there is no clear intruder according to these features. Let us focus on the other features that are not identical. For each such index j, we can compute the 4 values $H_a(a_j, b_j, c_j, d_j)$, $H_b(a_j, b_j, c_j, d_j)$, $H_c(a_j, b_j, c_j, d_j)$ and $H_d(a_j, b_j, c_j, d_j)$. We know that they cannot be all equal (or close to) 1 since their conjunction is not satisfiable: in fact, exactly one proportion has to be close to 0, thus spotting out the intruder for that component. Applying again a majority vote, we shall consider as global intruder the one which is intruder for the maximum number of components.

In the case where we have to 'Find the odd one out' among more than 4 items, diverse options are available. We may consider all the subsets of 4 items. For each such subset, we apply the previous method to exhibit an intruder (if any). Then the global intruder will be the one which is intruder for the maximum number of subsets.

7 CONCLUSION

The Boolean modeling of logical proportions which relate 4 items in terms of similarity and dissimilarity, and which may be viewed as a counterpart to numerical proportions, has led to identify a set of 120 distinct proportions. All these logical proportions have the same type of truth table, namely they are true for exactly 6 valuations (and thus false for the 10 remaining valuations). Among this set, only 8 proportions satisfy a so-called code-independency property which makes sure that the evaluation of the proportion remains unchanged when the truth values of the 4 components are changed into their complement (1 is changed into 0, and 0 into 1). This property is important since it ensures that the evaluation of logical proportions will not depend on the positive or negative encoding of the features of the considered items. This set of 8 remarkable proportions divides into 4 homogeneous proportions, and the 4 heterogeneous proportions. These two subsets can be strongly contrasted and appear to be complementary. The 6 valuation patterns that make true homogeneous proportions have all an even number of 1 (and consequently of 0), while for heterogeneous proportions the numbers are odd. Homogeneous proportions are symmetrical, while heterogeneous ones are not. Both types of proportions satisfy remarkable permutation properties. Interestingly enough, these two subsets of logical proportions can be related to two types of IQ tests or quizzes respectively of the type “Find the missing item” and of the type “Find the odd one out”. Thus, both from a formal viewpoint and from an applicative viewpoint, heterogeneous proportions appear as a perfect dual of the homogeneous ones. Ultimately, logical proportions provide an elegant unique framework for dealing with IQ tests, from Raven progressive matrices to Find the odd one out quizzes, in a uniform way. Generally speaking, beyond these illustrations, logical proportions still constitute an intriguing set of quaternary connectives, including diverse subsets with remarkable properties, that look interesting for different reasoning purposes where the ideas of similarity and dissimilarity play a role.

BIBLIOGRAPHY

[Abraham *et al.*, 2009] M. Abraham, D. M. Gabbay, and U. J. Schild. Analysis of the Talmudic argumentum a fortiori inference rule (kal vachomer) using matrix abduction.

Studia Logica, 92(3):281–364, 2009.

[Ciucci and Dubois, 2012] D. Ciucci and D. Dubois. Relationships between connectives in three-valued logics. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, editors, *Advances on Computational Intelligence - Proc. 14th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'12), Catania, July 9-13, Part I*, volume 297 of *Communications in Computer and Information Science*, pages 633–642. Springer, 2012.

[Correa *et al.*, 2012] W. Correa, H. Prade, and G. Richard. When intelligence is just a matter of copying. In L. De Raedt et al., editor, *Proc. 20th Europ. Conf. on Artificial Intelligence, Montpellier, Aug. 27-31*, pages 276–281. IOS Press, 2012.

[De Finetti, 1936] B. De Finetti. La logique des probabilités. In *Congrès International de Philosophie Scientifique*, pages 1–9, Paris, France, 1936. Hermann et Cie.

[Dubois and Prade, 1994] D. Dubois and H. Prade. Conditional objects as nonmonotonic consequence relationships. *IEEE Trans. on Systems, Man and Cybernetics*, 24:1724–1740, 1994.

[French, 2002] R. M. French. The computational modeling of analogy-making. *Trends in Cognitive Sciences*, 6(5):200 – 205, 2002.

[Gentner, 1983] D. Gentner. Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155–170, 1983.

[Hampshire, 2010] A. Hampshire. The odd one out tests of intelligence. http://www.cambridgebrainsciences.com/browse/reasoning/test/oddoneout, 2010.

[Klein, 1982] S. Klein. Culture, mysticism & social structure and the calculation of behavior. In *Proc. 5th Europ. Conf. in Artificial Intelligence (ECAI'82), Paris*, pages 141–146, 1982.

[Klein, 1983] S. Klein. Analogy and mysticism and the structure of culture (and Comments & Reply). *Current Anthropology*, 24 (2):151–180, 1983.

[Lovett *et al.*, 2010] A. Lovett, K. Forbus, and J. Usher. A structure-mapping model of Raven's progressive matrices. In *Proc. of the 32nd Annual Conference of the Cognitive Science Society, Portland, OR*, 2010.

[McGreggor and Goel, 2011] K. McGreggor and A. K. Goel. Finding the odd one out: a fractal analogical approach. In *Proc. 8th ACM Conf. on Creativity and Cognition*, pages 289–298, 2011.

[Miclet and Prade, 2009] L. Miclet and H. Prade. Handling analogical proportions in classical logic and fuzzy logics settings. In *Proc. 10th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09), Verona*, pages 638–650. Springer, LNCS 5590, 2009.

[Miclet *et al.*, 2008] L. Miclet, S. Bayoudh, and A. Delhay. Analogical dissimilarity: definition, algorithms and two experiments in machine learning. *JAIR, 32*, pages 793–824, 2008.

[Piaget, 1953] J. Piaget. *Logic and Psychology*. Manchester Univ. Press, 1953.

[Prade and Richard, 2009] H. Prade and G. Richard. Analogy, paralogy and reverse analogy: Postulates and inferences. In *Proc. 32nd Ann. Conf. on Artif. Intellig. (KI 2009), Paderborn*, volume LNAI 5803, pages 306–314. Springer, 2009.

[Prade and Richard, 2010a] H. Prade and G. Richard. Logical proportions - typology and roadmap. In E. Hüllermeier, R. Kruse, and F. Hoffmann, editors, *Computational Intelligence for Knowledge-Based Systems Design: Proc. 13th Inter. Conf. on Information Processing and Management of Uncertainty (IPMU'10), Dortmund, June 28 - July 2*, volume 6178 of *LNCS*, pages 757–767. Springer, 2010.

[Prade and Richard, 2010b] H. Prade and G. Richard. Multiple-valued logic interpretations of analogical, reverse analogical, and paralogical proportions. In *Proc. 40th IEEE Inter. Symp. on Multiple-Valued Logic (ISMVL'10), Barcelona, May*, pages 258–263, 2010.

[Prade and Richard, 2010c] H. Prade and G. Richard. Reasoning with logical proportions. In *Proc. 12th Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'10), Toronto, Ontario, Canada, May 9-13, (F. Z. Lin, U. Sattler, M. Truszczynski, eds.)*, pages 545–555. AAAI Press, 2010.

[Prade and Richard, 2011] H. Prade and G. Richard. Analogy-making for solving IQ tests: A logical view. In *Proc. 19th Inter. Conf. on Case-Based Reasoning, Greenwich, London, 12-15 Sept.*, LNCS, pages 561–566. Springer, 2011.

[Prade and Richard, 2012a] H. Prade and G. Richard. From analogical proportion to logical proportions. IRIT Tech. Rep., Univ. P. Sabatier, Toulouse,, 2012.

[Prade and Richard, 2012b] H. Prade and G. Richard. Homogeneous logical proportions: Their uniqueness and their role in similarity-based prediction. In G. Brewka, T. Eiter, and S. A. McIlraith, editors, *Proc. 13th Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'12), Roma, June 10-14*, pages 402–412. AAAI Press, 2012.

[Prade *et al.*, 2012] H. Prade, G. Richard, and B. Yao. Enforcing regularity by means of analogy-related proportions-a new approach to classification. *International Journal of Computer Information Systems and Industrial Management Applications*, 4:648–658, 2012.

[Raven, 2000] J. Raven. The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1 – 48, 2000.

[Schockaert and Prade, 2011] S. Schockaert and H. Prade. Interpolation and extrapolation in conceptual spaces: A case study in the music domain. In *Proc. 5th Inter. Conf. on Web Reasoning and Rule Systems (RR'11), Galway, Ireland, Aug. 29-30.* Springer, LNCS, 2011.

[Tversky, 1977] A. Tversky. Features of similarity. In *Psychological Review*, number 84, pages 327–352. American Psychological Association, 1977.

RICHMOND H. THOMASON

THE FORMALIZATION OF PRATICAL REASONING: PROBLEMS AND PROSPECTS

INTRODUCTION

Deontic logic, as traditionally conceived, provides only a deductive theory that constrains the states or possible worlds within which an agent should try to remain. As such, it only encompasses a small part of practical reasoning, which in general is concerned with selecting, committing to, and executing plans. In this article I try to frame the general challenge that is presented to logical theory by the problem of formalizing practical reasoning, and to survey the existing resources that might contribute to the development of such a formalization. I conclude that, while a robust, adequate logic of practical reasoning is not yet in place, the materials for developing such a logic are now available.

1 THE CHALLENGE OF FORMALIZING PRACTICAL REASONING

Practical reasoning is deliberation. It is reasoning about what to do. We do it all the time. Any day in our life will provide us with hundreds of examples of thinking about what to do. But it has been remarkably difficult to produce a comprehensive, adequate theory of practical reasoning. Part of the difficulty is that the topic is studied by different disciplines, each of these has something important to contribute, and it is unusual to find a study of practical reasoning that brings all of these perspectives together.

I will begin by considering examples of practical reasoning. (I suspect that the range of examples is broader than many people might imagine.) I will then propose a rationale for classifying these examples, and canvass the disciplines that have something useful to say about the reasoning.

A more or less comprehensive inventory of examples will provide an idea of what an adequate account of practical reasoning might look like. In the remainder of the paper, I try to say something about the challenges that an approach that begins to do justice to the subject would have to address.

1.1 Some Examples

All too many published discussions of practical reasoning—even book-length discussions—cover only a very small part of the territory. For that reason, it's vital to begin with a broad range of examples.

D. M. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*,
Handbook of Philosophical Logic 18, https://doi.org/10.1007/978-3-319-97755-3_3

Example 1. Ordering a meal at a restaurant.

The deliberating agent sits down in a restaurant and is offered a menu. Here, the problem is deciding what to eat and drink. Suppose that the only relevant factors are price and preferences about food. Even for a moderately sized menu and wine list, the number of possible combinations is over $400,000$. It would be very unlikely for an ordinary human being to work out a total preference ordering for each option. In fact, even though the decision will probably involve weighing preferences about food and drink against preferences about cost, the reasoning might well produce a decision without appealing to a general rule for reconciling these preferences.

Example 2. Deciding what move to make in a chess game.

In chess, an individual action needs to be evaluated in the context of its continuations. There is no uncertainty about the current state or the immediate consequences of actions, but much uncertainty about moves that the opponent might make. The *search space* (i.e., the number of possible continuations) is enormous—on the order of 10^{43}. Determining the value of positions involves conflicting criteria (e.g. positional advantages versus numerical strength); these conflicts must be resolved in comparing the value of different positions. In tournament chess, deliberation time is limited. These somewhat artificial constraints combine to concentrate the reasoning on exploration of a search space. Perhaps because of this, the reasoning involved in chess has been intensively investigated by psychologists and computer scientists, and influenced the classical work on search algorithms in AI; see [Simon and Schaeffer, 1992].

Example 3. Savage's omelet.

In [Savage, 1972][pp. 13–15], Leonard Savage describes the problem as follows.

> Your wife has just broken five good eggs into a bowl when you come in and volunteer to finish making the omelet. A sixth egg, which for some reason must either be used for the omelet or wasted altogether, lies beside the bowl. You must decide what to do with the unbroken egg. . . . you must decide between three acts only, namely, to break it into the bowl containing the other five, to break it into a saucer for inspection, or to throw it away without inspection.

This problem involves preferences about the desired outcomes, as well as risk, in the form of a positive probability that the egg is spoiled. The problem is to infer preferences over actions. The outcomes are manifest and involve only a few variables, the preferences over them are evident, and the probabilities associating each action with an outcome can be easily estimated. In this case, the reasoning reduces to the calculation of an expected utility.

Example 4. Designing a house.

This example is less obviously practical; it is possible for an architect to design a house without thinking much about the actions that will go into building it, leaving this to the contractor.[1] However, an architect's design becomes the builder's goals, and I would maintain that inferring goals is a form of practical reasoning. The reasoning combines constraint satisfaction and optimization, where again conflicts between competing desiderata may need to be resolved. Any real-life architect will also use *case-based reasoning*, looking in a library of known designs for one that is relevant, and modifying a chosen example to suit the present purpose.

Example 5. Deciding how to get to the airport.

This is a planning problem; the agent a has an inventory of actions, knows their preconditions and effects, knows the relevant features of the current state, and has as its goal a state in which a is at the airport. In its simplest form, the problem is to find a sequence of actions that will transform the current state into a state that satisfies the goal. Planning, or means-end reasoning, is one of the most intensively studied forms of reasoning in AI. The earliest planning algorithms made many simplifying assumptions about the planning situation and the conditions that a satisfactory plan must meet; over the years, sophisticated planning algorithms have been developed that depend on fewer of these assumptions and so can be used in a variety of realistic settings.[2]

Example 6. Cracking an egg into a bowl.

This is a case in which most of of us do the action automatically, with hardly any conscious reasoning. Probably most people can't remember the circumstances under which they learned how to do it. But the activity is complex: there are many ways to get it wrong. This example was proposed as a benchmark problem in the formalization of common-sense reasoning. The literature on this problem shows that the reasoning is surprisingly complicated, and it presupposes much common-sense knowledge; see, for instance, [Shanahan, 1997a]. This example is different from the previous ones in that the solution to the reasoning problem is acted out; the reasoning must engage motor systems, and it depends on these systems for grasping and manipulating objects according to

[1]Of course, a good design has to take into account how to build a house, in order to make sure that the design is feasible.

[2]See, for instance, [Reiter, 2001]. For the airport problem in particular, see [Lifschitz *et al.*, 2000].

plan. For obvious reasons, Savage ignored this part of the omelet problem.

Example 7. Playing table tennis.

Unlike chess, table tennis is a game in which practical reasoning has to be *online*; engaged in complex, real-time activities involving the perceptual and motor systems. For a novice, the reasoning may be exhausted by the need to keep the ball in play; experts may be able to engage in tactical reasoning. But there is no time to spare for reflection; the reasoning needs to be thoroughly connected to the ongoing process of play.

Example 8. Playing soccer.

Soccer is like table tennis, but with the added dimension of teamwork and the need to recognize and execute plays. This task was selected as a benchmark problem in robotics, and has been extensively studied. See, for instance, [Visser and Burkhard, 2007; Ros *et al.*, 2009; Asada *et al.*, 1999].

Example 9. Typing a message.

Typing an email message, composing it as you go along, starts perhaps with a general idea of what to say. The reasoning that produced a rough idea of the content may have taken place reflectively, but once composition has begun, several reasoning processes are engaged simultaneously, and have to be coordinated. The general idea of what to say has to be packaged in linguistic form, and this form has to be rendered by motor actions at the keyboard. For a skilled typist composing a straightforward message, these complex, practical tasks are combined and executed very quickly, perhaps at the rate of 70 words per minute. For this to happen, the interface between high-level linguistic reasoning and motor skills has to be very robust.

Example 10. Factory scheduling.

The factory scheduler has to produce, say on a daily basis, a sequence of manufacturing operations for each order to be processed that day, and a schedule allocating times and machines to these operations. This problem is notorious for the difficulty of the reasoning; it involves horrible combinatorics, uncertainty, limited time for reflection, and the resolution of many conflicting desiderata. Among the goals cited by [Fox and Clarke, 1991] are (1) meeting order dates, (2) minimizing work-in-process time, (3) maximizing allocation of factory resources, and (4) minimizing disruption of shop activity.

Part of the interest of this example lies in the difference in scale between this problem and Savage's omelet problem. It is not clear that there is any way to construct a single, coherent utility function for the task, by reconciling the four desiderata mentioned above. Any reconciliation will leave some managers unhappy: salesmen will favor goal (1), and production managers will favor goals (2)–(3), perhaps giving different weights to these. Nor is it feasible to produce a global probability function for a system with so many interacting variables.

Example 11. Ordering dessert.

Let's return the restaurant of Example 1. The main course is over, and our agent is offered a dessert menu and the choice of whether to order dessert. On the one hand, there is a direct desire for dessert, perhaps even a craving. This alternative is colored with and motivated by emotion, even if the emotion is not overwhelming. But suppose that there is a contrary emotion. The agent is unhappy with being overweight and has determined to eat less, and may have told others at the table about the decision to undertake a diet. This creates a conflict, coloring the choice of dessert with negative associations, perhaps even shame. The chief difference between this conflict and those in Examples 2 and 4 is that this decision is emotionally "warm;" the outcome may be influenced by a craving and the presence of the desired object. (Perhaps this is why some restaurants invest in dessert trays.)

Example 12. An unanticipated elevator.

A man decides to visit his stockbroker in person, something he has never done. He takes a bus to a stop near the stockbroker's downtown address, gets off the bus, locates the building and enters it. He finds a bank of elevators, and sees that the stockbroker is on the 22nd floor. This man has a strong dislike for elevators, and is not feeling particularly energetic that day. He reconsiders his plan.

Example 13. A woman is working in her garden.

She becomes hot and tired, and decides to take a break. Or she hears the telephone ringing in her house, and decides to answer it. Or she sees smoke coming out of the window of her house, and runs for help.

Example 14. The wrath of Achilles.

In Book I of *The Iliad*, the hero Achilles is outraged and dishonored by his warlord Agamemnon, who insults him and declares that he will take back, in compensation for his own loss and Achilles' disrespectful behavior, the captive woman that Achilles had received as his war prize.

Homer goes on to describe Achilles' reaction. Achilles is headstrong, but his reaction is partly physical and partly intellectual: his heart pounds with rage, but instead of acting immediately he asks himself a question: should he draw his sword and kill the king? To explain his decision, the poet brings in a god: Athena, invisible to everyone else, seizes him by the hair and persuades him to give in and be patient.

For our purposes, we can suppose that Athena is a literary device. The outrage leads to a direct desire to kill, but instead of acting on it, Achilles realizes that it would be better to restrain himself.

Even though it is "hot"—strongly informed by emotion—reasoning intervenes here between the emotional shading of the alternatives and an ensuing resolution to act.

Example 15. Deciding what to say at a given point in a conversation.

Conversation provides many good examples of deliberative reasoning. Where there is conscious deliberation, it is likely to be devoted to content selection. But the reasoning that goes into deciding how to express a given content can be quite complex.

Certainly, any adequate theory of practical reasoning must at least be compatible with this broad range of cases. Better, it should be capable of saying something about the reasoning involved in all of them. Even better, there should be a single architecture for practical reasoning, capable of dealing with the entire range of reasoning phenomena.[3] No doubt, there are special-purpose cognitive modules (e.g., for managing perception, motor behavior, and some aspects of language). But in the absence of convincing, independent psychological evidence it would be perverse to formulate a theory of a special type of practical reasoning, such as preference generation, probability estimation, or means-end reasoning, and to postulate a "cognitive module" that performs just this reasoning. All these types of reasoning can be involved in the same practical problem situation, and interact strongly. This methodology would be likely to produce an *ad hoc* and piecemeal account of practical reasoning.

1.2 Towards a classification

The examples in the previous section suggest a set of features that can be used to classify specimens of deliberative reasoning.

1. Are only a few variables (e.g., desiderata, causal factors, initial conditions) involved in the decision?

2. Do conflicting preferences need to be resolved in making the decision?

3. Is the time available for deliberation small compared to the time needed for adequate reflection?

4. Is the deliberation immediate? That is, will the intentions that result from the deliberation be carried out immediately, or postponed for future execution?

[3]For the idea of a cognitive architecture, see [Newell, 1992].

5. Is the deliberation carried out in "real time" as part of an ongoing activity involving sensory and motor activities?

6. Does the reasoning have to interface closely with sensory and motor systems?

7. Is the activity part of a group or team?

8. Does the context provide a definite, relatively small set of actions, or is the set of actions open-ended?

9. Is there certainty about the objective factors that bear on the decision?

10. Is the associated risk small or great?

11. Is the goal of deliberation a single action, or a sequence of actions?

12. Is continuous time involved?

13. Is the deliberation colored with emotions?

14. Is the action habitual, or automatic and unreflective?

15. Is there conscious deliberation?

16. Are there existing plans in play to which the agent is committed or that already are in execution?

Many of the differences marked by these features are matters of degree, so that the boundaries between the types of reasoning that they demarcate are fluid. This strengthens the case for a general approach to the reasoning. There is nothing wrong with concentrating on a special case to see what can be learned from it. Chess and decision problems that, like Savage's omelet, involve a solution to the "small worlds problem"[4] provide good examples of cases where this methodology has paid off. But to concentrate on these cases without paying any attention to the broad spectrum of examples runs the risk of producing a theory that will not be contribute usefully to something more general.

1.3 Disciplines and approaches

Many different disciplines have something to say about practical reasoning. The main theoretical approaches belong to one of the five following areas.

1. Philosophy

[4]This is the problem of framing a decision problem, concentrating only on the factors that are relevant.

2. Logic
3. Psychology
4. Decision Theory and Game Theory
5. Artificial Intelligence

Of course, there is a good deal of overlap and mixing of these approaches: AI, for instance, is especially eclectic and has borrowed heavily from each of the other fields. But work in each area is colored by the typical problems and methods of the discipline, and—typically, at least— has a distinctive perspective that is inherited from the parent discipline.

The following discussion of these five approaches is primarily interested in what each has to contribute to the prospects for formalizing practical reasoning.

Philosophy

The topic of practical reasoning goes back to Aristotle. In the Twentieth Century there was a brief revival of philosophical interest in the topic of "practical inference." This coincided more or less with early work on deontic and imperative logic, and was carried out by a group of logically minded philosophers and a smaller group of philosophicaly minded logicians. It is a little difficult to distinguish philosophy from logic in this work; I will more or less arbitrarily classify Kenny and some others as philosophers for the purposes of this exposition, and von Wright as a logician.

Post-Fregean interest in imperative logic seems to have begun about the time of World War 2, with [Jørgensen, 1937-1938; Hofstadter and McKinsey, 1939; Ross, 1941]. Later, in the 1960s,[5] some British philosophers became interested in the topic. This period saw 10 or more articles relevant appearing in journals like *Analysis*. Of these, [Kenny, 1966] seems to have the most interesting things to say about the problem of formalizing practical reasoning.[6]

Kenny begins with Aristotle's practical syllogism, taking several specimens of means-end reasoning from the Aristotelian corpus, and beginning with the following example, based on a passage in *Metaphysics* 1032b19.

Example 16. A doctor prescribing.

This man is to be healed.
If his humors are balanced, he will be healed.
If he is heated, his humors will be balanced.

[5] Judging from internal evidence, the work of Richard Hare influenced this episode of interest in the topic. Elizabeth Anscombe [Anscombe, 1958] may also have been an influence, as well as G.H. von Wright.

[6] For more about this period, see [Green, 1997].

If he is rubbed, he will be heated.
So I'll rub him.

The premisses of the reasoning, according to Kenny, are either (i) desires or duties, or (ii) relevant facts. And he characterizes the conclusion as an action.[7] Kenny points out that this sort of reasoning doesn't fit Aristotelian syllogistic, and that a straightforward modern formalization of it would be invalid. To put it crudely, the inference from P, $Q \rightarrow P$, $R \rightarrow Q$, and $S \rightarrow R$ to S is invalid.

Here, I think Kenny has indicated an important type of practical reasoning, and pointed out a glaring problem with the propositional calculus as a formalization medium. Unfortunately, the theory that he proposes in this paper doesn't seem to solve the problem of providing an account of validity that matches the reasoning. In fact, there are many glaring problems with the crude Propositional Calculus formalization of Example 16, involving the deductive formulation of the reasoning as well as the faithfulness of the formalization to the language of the example.

The failure of Kenny's proposal and of similar ones at the time seems to originate in a lack of logical resources that do justice to the problem. The Propositional Calculus is certainly not the right tool, and deduction is certainly not the right characterization of the reasoning. The only idea that was explored at the time was that of providing a logic of "imperative inference." This idea might help with one problem: formalizing the first premiss of Example 16, which does not seem like a straightforward declarative. But it can't begin to address the challenge posed by the invalidity of the argument. Besides, the idea of an imperative logic didn't lead to anything very new, because of another trend that was taking place at about the same time.

This trend, which tried to absorb imperative and practical inference into some sort of modal logic, was also underway in the 1960s. [Lemmon, 1965] provides a logic of imperatives that prefigures the STIT approach of [Belnap, Jr. *et al.*, 2001], hence a modal approach that brings in the idea of causing a state of affairs. And [Chellas, 1969], recommends and develops a reduction of imperative logic to a more standard deontic logic. This idea provides formal systems with excellent logical properties. But it does so at the expense of changing the subject and leaving the central problem unsolved. Reasoning in deontic logic is deductive, and if you formalize typical specimens of means-end reasoning like Example 16 in these systems, the formalizations will be invalid.

Even though the literature shows a sustained series of attempts in this

[7]The Aristotelian texts make it pretty clear that Aristotle considered the conclusion to be an action. But for our purposes, it would work better to think of the conclusion as an expression of intention. In some circumstances—when the deliberation is concerned with immediate action and the reasoning is sufficiently persuasive, there is no gap between intention and action.

period to formalize practical inference, the work didn't lead to anything like a consensus, and produced no sustainable line of logical development. In retrospect, we can identify several assumptions that rendered the formalization project unsustainable:

1. These philosophers relied too much on deductive inference, with the propositional calculus as a paradigm, and too little on models;

2. They tended to work with overly simple formal languages;

3. They didn't bring actions into the formalization explicitly;

4. They missed the insight that means-end reasoning is more like abduction or heuristic search than deduction.

As we will see in Section 1.3, more recent and quite separate developments in computer science have yielded sophisticated logics of means-end reasoning, effectively solving the formalization problem that led to an earlier philosophical impasse in the 1960's and 1970's. The moral seems to be that formalization projects of this sort can involve multiple challenges, and that it can be hard to address these challenges without a body of applications and a community of logicians committed to formalizing the applications and mechanizing the reasoning.

Meanwhile, philosophers seem to have drawn the conclusion that close attention to the reasoning, and searching for formalizations, is not likely to be productive. In the more recent philosophical work on practical reasoning, it is actually quite difficult to find anything that bears on the formalization problem. Almost entirely, the philosophical literature is devoted to topics that might serve to provide philosophical foundations for the theory of practical reasoning—if there were such a theory. Even if, as Elijah Millgram claims in [Millgram, 2001], the driving issue in the philosophy of practical reasoning is to determine which forms of practical reasoning are correct, philosophers seem to pursue this inquiry with informal and very loose ideas of the reasoning itself. In many cases—for instance, the issue of whether intentions cause actions—no formalization of the reasoning is needed for the philosophical purposes. In other cases, however, a formal theory of practical reasoning might help the philosophy, refining some old issues and suggesting new ones.

Even though some philosophers maintain positions that would sharply limit the scope of practical reasoning (reducing it, for instance, to means-end reasoning), I don't know of any explicit, sustained attempt in the philosophical literature to delineate what the scope of practical reasoning should be. I don't see how to do this without considering a broad range of examples, as I try to do above in Section 1.1. But in fact, examples of practical reasoning are thin on the ground in the philosophical literature; in [Millgram, 2001], for instance, I counted only 12 examples of practical reasoning in 479 pages—and many of these were brief illustrations of general points.

Logic

There are few departments of logic, and work in logic bearing on practical reasoning tends to be carried out in the context of either Philosophy or Computer Science, and to be influenced by the interests of the parent disciplines. There are, in fact, two separate strands of logical research, one associated with Philosophy and the other with Artificial Intelligence. These have interacted less than one might wish.

Philosophical logic. Georg Henrik von Wright was explicitly interested in practical reasoning, from both a philosophical and a logical standpoint. Most of his writings on the topic are collected together in [von Wright, 1983]; these were published between 1963 and 1982. Like Kenny, von Wright begins with Aristotle's practical syllogism. But he avoids the problem of invalidity by strengthening premisses that introduce ways of achieving something. Von Wright's version of Example 16 would look like this:

> I want to heal this man.
> Unless his humors are balanced, he will not be healed.
> Unless he is heated, his humors will not be balanced.
> Unless he is rubbed, he will not be heated.
> Therefore I must rub him.

By departing from Aristotle's formulation, von Wright makes it easier to formulate the inference in a deontic logic, and to see how the formalization might be valid. At the same time, he is making it more difficult to fit the formalization to naturally occurring reasoning. As in this example, where, for instance, there is surely more than one way to heat the patient, the means that a deliberator chooses in typical means-end reasoning will not be the only way to achieve the end.

This simplification makes it easier for von Wright to propose modal logic, and in particular deontic logic, as the formalization medium for practical reasoning. Von Wright also characterizes his version of deontic logic as a "logic of action," but all this seems to mean is that the atomic formulas of his language may formalize things of the form 'Agent A does action a.' He has little or nothing to say about reasoning about action.

I will not say much here about the subsequent history of deontic logic as a part of philosophical logic. As the field developed, it acquired its own problems and issues (such as the problem of reparational obligations), but as philosophers concentrated on declarative formalisms and deductive logic, the relevance to practical reasoning, and even means-end reasoning, that von Wright saw in his in early papers such as [von Wright, 1963], attenuated.

Although the subsequent history of deontic logic was less directly concerned with practical reasoning, it shows a healthy tendency to concentrate on naturally occurring problems that arise in reasoning about obligation. This work has a place in any general theory of practical reasoning. Obligations play a role as constraints on means-end reasoning, and reasoning about obligations has to be flexible to cope with changing circumstances.

Also, the problem of modeling conditional obligations has produced a large literature on the relationship between modal logic and preference.[8] Of course, reasoning about preferences intrudes into practical reasoning in many ways. How to fit it in is something I am not very clear about at the moment; part of the problem is that so many different fields study preferences, and preferences crop in so many different types of practical reasoning. Maybe the best thing would be to incorporate preferences in a piecemeal way, and hope that a more general and coherent approach might emerge from the pieces.

The STIT approach to agency was already mentioned in Section 1.3. This provides a model-theoretic account of how actions are related to consequences that is quite different from the ones that emerged from the attempts in AI to formalize planning. The connections of STIT theory to practical reasoning are tenuous, and I will not have much to say about it.

Philosophy and philosophical logic have served over the years as a source of ideas for extending the applications of logic and developing logics that are appropriate for the extensions. One would hope that philosophy would continue to play this role. But—at least, for areas of logic bearing on practical reasoning—the momentum has shifted to computer science, and especially to logicist AI and knowledge representation. This trend began around 1980, and has accelerated since. Because many talented logicians were attracted to computer science, and because the need to relate theories to working implementations provided motivation and guidance of a new kind, this change of venue was accompanied by dramatic logical developments, and improved insights into how logic fits into the broader picture. I would very much like to see philosophy continue to play its foundational and creative role in developing new applications of logic, but I don't see how this can happen in the area of practical reasoning unless philosophers study and assimilate the recent contributions of computer scientists.

The point is illustrated by [Gabbay and Woods, 2005]. The paper is rare among contemporary papers in urging the potential importance of a logic of practical reasoning, but—in over 100 pages—it is unable to say what a coherent, sustained research program on the topic might be like. It does mention some important ideas, such as taking the agent into account, as well

[8] See, for instance, [Hansson, 2001; Jones and Carmo, 2002].

as nonmonotonic and abductive reasoning, but offers no explicit, articulated theories and in fact is hesitant as to whether logic has a useful role to play, repeating some doubts on this point that have been expressed by some roboticists and cognitive psychologists. Although it cites a few papers from the AI literature, the citations are incidental; work on agent architectures, abductive reasoning, and means-end reasoning goes unnoticed. Part of the problem is that the authors seem to feel that work in "informal logic" might be useful in approaching the problem of practical reasoning—but the ideas of informal logic are too weak to provide any helpful guidance. If we are interested in accounting for the practical reasoning of agents, we have to include computer programs. For this, we need formal logic—but formal logic that is applicable.

I couldn't agree more with Gabbay and Woods that logicians should be concerned with practical reasoning. But to make progress in this area, we need to build on the accomplishments of the formal AI community.

Psychology

From the beginning of cognitive psychology, a great deal of labor has gone into collecting protocols from subjects directly engaged in problem-solving, much of it practical. Herbert Simon and Allen Newell were early and persistent practitioners of this methodology. This material contains many useful examples; in fact, it helped to inspire early characterizations of means-end reasoning in Artificial Intelligence.

As early as 1947, in [Simon, 1947], Simon had noted divergences between decision-making in organizations and the demands of ideal rationality that are incorporated in decision theory; he elaborated the point in later work. An important later trend that began in psychology, with the work of Amos Tversky and Daniel Kahneman, studies these differences in more detail, providing many generalizations about the way people in fact make decisions and some theoretical models; see, for instance, [Kahneman and Tversky, 1979; Tversky and Kahneman, 1981].

Tversky and Kahneman's experimental results turned up divergences between ideal and actual choice-making that were not obviously due, as Simon had suggested, merely to the application of limited cognitive resources to complex, time-constrained problems. Since their pioneering work, this has become a theme in later research.

All this raises a challenging foundational problem, one that philosophers might be able to help with, if they gave it serious attention. What level of idealization is appropriate in a theory of deliberation? What is the role of "rationality" in this sort of idealization? Is there a unique sort of rationality for all practically deliberating agents, or are there many equally reasonable ways of deliberation, depending on the cognitive organization and deliberative style of the agent? Is the notion of rationality of any use at all, outside the range of a very limited and highly idealized set of decision problems? Probably it would be unwise to address these problems before attempting to provide a more adequate formalization of practical reasoning—that would be likely to delay work on the formalization indefinitely. But the problems are there.

Nowadays, the cognitive psychology of decision-making has migrated into Economics and Management Science, and is more likely to be found in economics departments and schools of business than in psychology departments. This doesn't affect the research methods much, but it does improve the lines of communication between researchers in behavioral economics and core areas of economics. As a result, economic theorists are becoming more willing to entertain alternatives to the traditional theories.

Decision Theory and Game Theory

The literature in these areas, of course is enormous, and most of it has to do with practical reasoning. But traditional work in game theory and decision theory concentrates on problems that can be formulated in an idealized form—a form in which the reasoning can be reduced to deriving an optimum result by calculation.[9] As a result, work in this tradition tends to neglect much of the reasoning in practical reasoning. Of course, an agent must reason to wrestle a practical problem into the required form—to solve Savage's "small worlds problem"—but the literature in economics tends to assume that somehow the problem has been framed, without saying much if anything about the reasoning that might have gone into this process. (Work in decision analysis, of course, is the exception.) And once a problem has been stated in a form that can be solved by calculation, there is little point in talking about deliberative processes.

If we are concerned with the entire range of examples presented in Section 1.1, however, we find many naturally occurring problems that don't fit this pattern; and some of these, at least, exhibit discursive, inferential reasoning. This is one reason why I believe that a general theory of practical reasoning will reserve an important place for qualitative reasoning, and especially for inferential reasoning—the sort of reasoning that gives formalization and logic a foothold. In this respect, Aristotle was on the right track.

[9]Microeconomists and statisticians are not the only ones who have taken this quantitative, calculational paradigm to heart. Many philosophers have accepted the paradigm as a model of practical reasoning and rationality. See, for instance, [Skyrms, 1990], a book-length study of practical deliberation, which takes the only relevant theoretical paradigms to be decision theory and game theory, and takes them pretty much in the classical form. Skyrms' book and the many other philosophical studies along these lines have useful things to say; my only problem with this literature is the pervasive assumption that practical reasoning can be comprehensively explained by quantitative theories based on the assumption that agents have global probability and utility functions.

At the very least, practical reasoning can involve inference and heuristic search, as well as calculation. (Calculation, of course, is a form of reasoning, but is not inferential, in the sense that I intend.) Any theory of practical reasoning that emphasizes one sort of reasoning at the expense of others must sacrifice generality, confining itself to only a small part of the territory that needs to be covered by an adequate approach. The imperialism of some of those (mainly philosophers, these days) who believe that there is nothing to rationality or practical reasoning other than calculations involving probability and utility, can partly be excused by the scarcity of theoretical alternatives. I will argue in this article that the field of Artificial Intelligence has provided the materials for developing such alternatives.

As I said in Section 1.3, research in behavioral economics has made microeconomists generally aware that, in their original and extreme form, the idealizations of decision theory don't account well for a broad range of naturally occuring instances of practical reasoning. Attempts to mechanize decision-making led computer scientists to much the same conclusion.

A natural way to address this problem begins with decision theory in its classical form and attempts to relax the idealizations. Simon made some early suggestions along these lines; other, quite different proposals. can be found in [Weirich, 2004] and [Russell and Wefald, 1991]. And other relaxations of decision theory have emerged in Artificial Intelligence: see the discussion of Conditional Preference Nets below, in Section 1.3. Still other relaxations have emerged out of behavioral economics, such as Tversky and Kahneman's Prospect Theory; see [Kahneman and Tversky, 1979].

Programs of this sort are perfectly compatible with what I will propose here. A general account of practical reasoning has to include calculations that somehow combine probability (represented somehow) and utility (represented somehow), in order to estimate risk. The more adaptable these methods of calculation are to a broad range of realistic cases, the better. I do want to insist, however, that projects along these lines can only be part of the story. Anyone who has monitored their own decision making must be aware that not all practical reasoning is a matter of numerical calculation; some of it is discursive and inferential. A theory that does justice to practical deliberation has to include both forms of reasoning. From this point of view, the trends from within economics that aim at practicalizing game theory and decision theory are good news. From another direction, work in Artificial Intelligence that seeks to incorporate decision theory and game theory into means-end reasoning is equally good news.[10]

In many cases of practical reasoning, conflicts need to be identified and removed or resolved. Work by economists on value tradeoffs is relevant and useful here; the classical reference is [Keeney and Raiffa, 1976], which

[10] For a survey, now getting rather old, see [Blythe, 1999]. For an example of a more recent, more technical paper, see [Sanner and Boutilier, 2009].

contains analyses of many naturally occuring examples.

Computer science and artificial intelligence

For most of its existence, the field of AI has been concerned with realistic decision problems, and compelled to formalize them. As the field matured, the AI community looked beyond procedural formalizations in the form of programs to declarative formulations and logical theories. Often AI researchers have had to create their own logics for this purpose.[11] Here, I will be concerned with three trends in this work: those that I think can offer most to the formalization of practical reasoning. These three are: means-end reasoning, reasoning about preferences, and agent architectures.

Dynamic logic and imperative inference. When an agent is given instructions and intends to carry them out unquestioningly, there is still reasoning to be done, and the reasoning is practical[12]—although, as the instructions become more explicit, the less scope there is for interesting reasoning from the human standpoint. Even so, the case of computer programs, where explicitness has to be carried out ruthlessly, can be instructive, because it shows how logical theory can be useful, even when the reasoning paradigm is not deductive.

A computer program is a (possibly very large and complex) imperative, a detailed instruction for carrying out a task. Many of its components, such as

$$\textbf{let } y \textbf{ be } x$$

("set the value of y to the current value of x") are imperatives, although some components, like the antecedent of the conditional instruction

$$\textbf{if } (x < y \textbf{ and } \textbf{not}(x = 0)) \textbf{ then let } z \textbf{ be } y/x$$

are declarative.

Inference, in the form of proofs or a model theoretic logical consequence relation, plays a small part in the theory of dynamic logic. Instead, *execution* is crucial. This idea is realized as the series of states that the agent (an idealized computer) goes though when, starting in a given initial state, it executes a program. Because states can be identified with assignments to variables, there are close connections to the familiar semantics of first-order logic.

Dynamic logic is useful because of its connection to *program verification.* A program specification is a condition on what state the agent will reach if it

[11]Throughout his career, John McCarthy was a strong advocate of this approach, and has done much of the most important work himself. See [McCarthy and Hayes, 1969] for an early statement of the methodology, and a highly influential proposal about how to formalize means-end reasoning.

[12]See [Lewis, 1979].

executes the program; if the initial state of a parsing program for an English grammar G, for instance, describes a string of English words, the program execution should eventually halt. Furthermore, (1) if the string is grammatical according to G, the executor should reach a final state that describes a parse of the string, and (2) if the string is not grammatical according to G it should reach a final state that records its ungrammaticality.

Dynamic logic has led to useful applications and has made important and influential contributions to logical theory. It is instructive to compare this to the relatively sterile philosophical debate concerning "imperative inference" that took place in the 1960s and early 1970s.[13] To a certain extent, the interests of the philosophers who debated imperative inference and the logicians who developed dynamic logic were different. Among other things, the philosophers were interested in applications to metaethics, and computational applications and examples didn't occur to them.

But the differences between philosophers and theoretical computer scientists, I think are relatively unimportant; some of the philosophers involved in the earlier debate were good logicians, and would have recognized a worthwhile logical project if it had occurred to them. In retrospect, three factors seem to have rendered the earlier debate unproductive:

1. Too great a reliance on deductive paradigms of reasoning;

2. Leaving a model of the executing agent out of the theoretical picture;

3. onfining attention to simple examples.

In dynamic logic, the crucial semantic notion is the correctness of an imperative with respect to a specification. Logically interesting examples of correctness are not likely to present themselves without a formalized language that allows complex imperatives to be constructed, and without examples of imperatives that are more complicated than 'Close the door'. (The first example that is presented in [Harel *et al.*, 2000] is a program for computing the greatest common divisor of two integers; the program uses a *while*-loop.) And, of course, a model of the executing agent is essential to the logical theory. In fact, what is surprising is how much logic can be accomplished with such a simple and logically conservative agent model.

As I said, the activity of interpreting and slavishly executing totally explicit instructions is a pretty trivial form of practical reasoning. But a logic of this activity is at least a start. I want to suggest that, in seeking to formalize practical reasoning, we should be mindful of these reasons for the success of dynamic logic, seeking to preserve and develop them as we investigate more complex forms of practical reasoning.

---

[13]See, for instance, [Williams, 1963; Geach, 1963] as well as [Kenny, 1966], which was discussed above, in Section 1.3.

Planning and the formalization of means-end reasoning. Perhaps the most important contribution of AI to practical reasoning is the formalization of means-end reasoning, along with appropriate logics, and an impressive body of research into the metamathematical properties of these logics, and their implementation in planning systems.[14]

This approach to means-ends reasoning sees a planning problem as consisting of the following components:

1. An initial state. (This might be described by a set of literals, or positive and negative atomic formulas.)

2. Desiderata or goals. (These might consist of a set of formulas with one free variable; a state that satisfies these formulas is a goal state.)

3. A set of actions or operators. Each action a is associated with a causal axiom, saying that if a state s satisfies certain preconditions, then a state $\textsc{Result}(a, s)$ that results from performing a in s will satisfy certain postconditions.

Here, the fundamental logical problem is how to define the state or set of successor states[15] resulting from the performance of an action in a state. (Clearly, not all states satisfying the postconditions of the action will qualify, since many truths will carry over to the result by "causal inertia.") This large and challenging problem spawned a number of subproblems, of which the best-known (and most widely misunderstood) is the *frame problem.* Although no single theory has emerged from years of work on this problem as a clear winner, the ones that have survived are highly sophisticated formalisms that not only give intuitively correct results over a wide range of test cases, but provide useful insights into reasoning about actions. Especially when generalized to take into account more realistic circumstances, such as uncertainty about the current state and concurrency or nondeterminism, these planning formalisms deliver logical treatments of means-end reasoning that go quite far towards solving the formalization problem for this part of practical reasoning.

I will try to say more about how these developments might contribute to the general problem of formalizing practical reasoning below, in Section 2.3.

Reasoning about preferences. It is hard to find AI applications that don't involve making choices. In many cases, it's important to align these choices with the designer's or a user's preferences. Implementing such

[14][Allen *et al.*, 1990] is a collection of early papers in the field. Both [Shanahan, 1997b] and [Reiter, 2001] describe the earlier logical frameworks and their later generalizations; [Reiter, 2001] also discusses implementation issues.

[15]Depending on whether we are working with the deterministic or the nondeterministic case.

preference-informed choices requires (i) a representation framework for preferences, (ii) an elicitation method that yields a rich enough body of preferences to guide the choices that need to be made, and (iii) a way of incorporating the preferences into the original algorithm.

Any attempt to extract the utilities needed for even a moderately complex, realistic decision problem will provide motives for relaxing the classical economic models of utility; but the need for workable algorithms seems to sharpen these motives. See [Goldsmith and Junker, 2008] for examples and details, and [Doyle, 2004], which provides a wide-ranging foundational discussion of the issues, with many references to the economics literature.

Of the relaxations of preference that have emerged in AI, *Ceteris Paribus* Preference Nets are one of the most widely used formalisms.[16] As in multi-attribute utility theory, the outcomes to be evaluated are characterized by a set of features. A parent-child relation must be elicited from a human subject; this produces a graph called a *CP-net*. The parents of a child feature are the features that directly influence preferences about the child. For instance, the price of wheat in the fall (high or low) might influence a farmer's preferences about whether to plant wheat in the spring. If the price will be high, the farmer prefers to plant wheat; otherwise, he prefers not to plant it. On the other hand, suppose that in the farmer's CP-net the price of lumber is unrelated to planting wheat. It can then be assumed that preferences about planting wheat are independent of the price of lumber.

To complete the CP-net, a preference ranking over the values of a child feature must be elicited for each assignment of values to each of the parent features.

Acyclic CP-nets support a variety of reasoning applications (including optimization), and—combined with means-end reasoning—provide an approach to preference-based planning.[17] And in many realistic cases it is possible to extract the information needed to construct a CP-net.

There are extensions of this formalism that allow for a limited amount of reasoning about the priorities of features in determining overall preferences; see [Brafman *et al.*, 2006].

The work in AI on preferences, like decision analysis, tends to concentrate on extracting preferences from a user or customer. Practical reasoning, however, produces a different emphasis. Some of the examples in Section 1.1—for instance, Examples 1, 4, 10, and 11—were designed to show that preferences are not automatically produced by the environment, by other agents, by the emotions, or by a combination of these things. We deliberate about what is better than what, and preferences can be the outcome of practical reasoning.[18] The status of an agent trying to work out its own preferences, and of a systems designer or decision analyst trying to

[16] See, for instance, [Domschlak, 2002; Boutilier *et al.*, 2003].

[17] See [Baier and McIlraith, 2008] for details and further references.

[18] For some preliminary and sketchy thoughts about this, see [Thomason, 2002].

work out the preferences of a person or an organization, may be similar in some ways, but I don't think we can hope that they are entirely the same. Nevertheless, insights into methods for extracting preferences from others might be helpful in thinking about how we extract our own preferences.

Agent architectures. A nonexecuting planning agent is given high-level goals by a user, as well as the declarative information about actions and the current state of things, as well perhaps as preferences to be applied to the planning process. With this information, it performs means-end reasoning and passes the result along to the user in the form of a plan.

This agent is not so different from the simple instruction-following agent postulated by dynamic logic; its capabilities are limited to the execution of a planning program, and it has little or no autonomy. But—especially in time-limited planning tasks—it may be difficult to formulate a specification, because the notion of what counts as an optimal plan in these condition is unclear.

When the planning agent is equipped with means of gathering its own information, perhaps by means of sensors, and is capable of performing its own actions, the situation is more complicated, and more interesting. Now the agent is interacting directly with its environment, and not only produces a plan, but must adopt it and put it in to action. This has a number of important consequences. The agent will need to perform a variety of cognitive functions, and to interleave cognitive performances with actions and experiences.

1. Many of the agent's original goals may be conditional, and these goals may be activated by new information received from sensors. This is not full autonomy, but it does provide for new goals that do not come from a second party.

2. Some of these new goals may be urgent; so the agent will need to be interruptable.

3. It must commit to plans—that is, it must form intentions. These intentions will constrain subsequent means-end reasoning, since conflicts between its intentions and new plans will need to be identified and eliminated.

4. It will need to schedule the plans to which it has committed.

5. It will need to monitor the execution of its plans, to identify flaws and obstacles, and repair them.

Recognizing such needs, some members of the AI community turned their attention from inactive planners to *agent architectures*, capable of integrating some of these functions. Early and influential work on agent architectures was presented in [Bratman *et al.*, 1988]; this work stressed the importance of intentions, and the role that they play in constraining future planning.

Any means-end reasoner needs desires (in the form of goals) and beliefs (about the state of the world and the consequences of actions). As Bratman, Israel, and Pollock point out, an agent that is implementing its own plans also needs to have intentions. Because of the importance of these three attitudes in the work that was influenced by these ideas, architectures of this sort are often known as *BDI architectures*. For an extended discussion of BDI architectures, with references to the literature up to 2000, see [Wooldridge, 2000]. See also [Georgeff *et al.*, 1999].

Work in "cognitive robotics" provides a closely related, but somewhat different approach to agent architectures. Ray Reiter, a leading figure in this area, developed methods for integrating logical analysis with a high-level programming language called GoLog, an extension of Prolog. Reiter's work is continued by the Cognitive Robotics Group at the University of Toronto.

Developments in philosophical logic and formal semantics have provided logics and models for propositional attitudes; for instance, see [Fagin *et al.*, 1995; Fitting, 2009]. Using these techniques, it is possible to formulate a metatheory for BDI agency. Such a metatheory is not the architecture; the reasoning modules of a BDI agent and overall control of reasoning has to be described procedurally. But the metatheory can provide specifications for some of the important reasoning tasks. Wooldridge's logic of rational agents, $\mathcal{LORA}$, develops this idea; see [Wooldridge, 2000].

A final word. Logicist AI has struggled to maintain a useful relation to applications, in the form of workable technology. Although the struggle has been difficult, many impressive success stories have emerged from this work—enough to convince the larger AI community of the potential value of this approach. The incentive to develop working applications has, I believe, been very helpful for logic, enabling new ideas that would not have been possible without the challenges posed by complex, realistic reasoning tasks.

Practical reasoning is not quite the same as logicist AI, or even the logical theory of BDI agents. But the successful use of logical techniques in this area of AI provides encouragement for a logical approach to practical reasoning. And, of course, it provides a model for how to proceed.

2 TOWARDS A FORMALIZATION

The challenge is this: how to bring logical techniques to bear on practical reasoning, and how to do this in a way that is illuminating, explanatory, and useful? In the rest of this article, I will only try to provide an agenda for addressing this challenge. The agenda divides naturally into subprojects.

Some of these subprojects can draw on existing work, and especially on work in AI, and we can think of them as well underway or even almost completed. Others are hardly begun.

2.1 Relaxing the demands of formalization

Let's return to the division between theoretical and practical reasoning.

Traditionally, domains that involve theoretical reasoning are formalized using what Alonzo Church called the "logistic method."[19] This method aims to formulate a formal language with an explicit syntax, a model-theoretically characterized consequence relation, and perhaps a proof procedure. Traditional formalizations did not include a model of the reasoning agent, except perhaps, in the highly abstract form of a Turing machine—this sort of agent is guaranteed whenever the consequence relation is recursively enumerable.

When it comes to practical reasoning, I believe that we have to be prepared to relax Church's picture of logical method.[20]

My own proposal for a relaxation is this: (1) we need to add a model of the reasoning agent, (2) we need to identify different phases of practical reasoning in agent deliberation, and different ways in which logic might be involved in each phase of the reasoning, and (3) consequently, we need to be prepared to have a logical treatment that is more pluralistic and less unified.

2.2 Agent architectures and division of logical labor

How should we model an agent that is faced with practical reasoning problems? In Section 1.1, I suggested that we should aim at, or at least acknowledge the existence of, a very broad range of reasoning problems. Suppose, for instance, that we classify the types of reasoning that we may need to consider in terms of the sort of conclusion that is reached. In view of the examples that were presented in Section 1.1, we will need to be prepared for the agent to infer:

1. Goals, which then invoke planning processes;

2. Plans, and the subgoals or means that emerge from plans;

3. Preferences emerging from reasoning about tradeoffs and risk;

---

[19] [Church, 1959][pp. 47–58].

[20] In fact, writing in 1956, Church was uncomfortable with semantics and model theory. He included these topics, but in a whisper, using small type. Over 50 years later, we have become quite comfortable with model theory and semantics, and are more likely to insist on this ingredient than on proof procedures. And in areas where logic is applied, we have become increasingly willing to bring the reasoning agent into the picture.

4. Intentions, commitments about what to do, and (to an extent) about when to do it;

5. Immediate decisions about what plan to execute;

6. Immediate, engaged adjustments of ongoing activities and plan executions, and shifts of attention that can affect the task at hand.

The examples in Section 1.1 were chosen, in part, to illustrate these activities. These sorts of deliberation are distinct, and all are practical. Although some of them can be automatic, they all can involve deliberate reasoning.

These six activities comprise my (provisional) division of practical reasoning into subtasks, and of the deliberating agent into subsystems. Each of them provides opportunities for logical analysis and formalization. I will discuss them in turn.

2.3 Means-end reasoning

This is the best developed of the six areas. We can refer to the extensive AI literature on planning and means-end reasoning not only for well developed logical theories, but for ideas about how this deliberative function interacts with the products of other deliberative subsystems—for instance, with preferences, and with plan monitoring and execution.

2.4 The practicalization of desires

On the other hand, work in AI on means-end reasoning, and on BDI agents, has little or nothing to say about the emotions and the origins of desires. In general, it is assumed that these come from a user—although the goals may be conditional, so that they are only activated in the appropriate circumstances. In principle, there is no reason why goals couldn't be inferred or learned. But the relevant reasoning processes have not, as far as I know, been formalized.

In truly autonomous agents some desires—perhaps all—originate in the emotions. Although a great deal has been written about the emotions, it is hard to find work that could serve a useful logical purpose.[21]

However desires originate, although they may be emotionally colored, they may not all be emotionally "hot." And to be useful in reasoning, some desires must be conditional, and self-knowledge about conditional desires must be robust. My preference for white wine this evening will probably

[21]Not [Solomon, 1976], which has a chapter on "Reason and the passions," a section on "The Rationality of the emotions," and a chapter on "The logic of the emotions." Not [Minsky, 2006], written by an author who knows something about AI. But work on modeling artificial characters for applications in areas like interactive fiction might be useful; see [Bates, 1994].

be accompanied by feelings of pleasure when I think about the refreshing taste of white wine. But the feeling of hypothetical pleasure is relatively mild; I am certainly not carried away by the feeling. AI systems builders are interested in obtaining a large body of conditional preferences from users because preferences need to be brought to bear under many different circumstances. Therefore a user's unconditional preferences—the preferences that are activated in the actual state of affairs—will not in themselves be very useful. Fully autonomous agents need conditional preferences as well, in planning future actions and in contingency planning.

Perhaps—to develop the example of preference for white wine a bit further—the only mechanism that is needed to generate conditional desires is the ability to imagine different circumstances, together with the ability to color these circumstances as pleasant (to some degree), and unpleasant (to some degree). But it is unlikely to be this simple, because pleasantness is not monotonic with respect to information: I find the idea of a glass of white wine quite pleasant, but the idea of a glass of white wine with a dead fly in it quite unpleasant. Also, my feelings about some imagined situations can be mixed, with elements that I find pleasant and elements that I find unpleasant. At this point, I might have to invoke a conflict resolution method that has little or nothing to do with the emotions.

This leads to a further point: there is a difference between raw or immediate desires, or *wishes*, and all-things-considered desires, or *wants*. This is because desires can not only conflict with one another, but with beliefs. And, when they conflict with beliefs, desires must be overridden: to do otherwise would be to indulge in wishful thinking.

In [Thomason, 2000], I explored the possibility of using a nonmonotonic logic to formalize this sort of practicalization of desires. The target reasoning consisted of deliberations such as the following. (The deliberator is a hiker who forgot her rain gear.)

1. I think it's going to rain.
2. If it rains, I'll get wet.
3. If I get wet, I'll stay wet unless I give up and go home.
4. I wouldn't like to stay wet.
5. I wouldn't like to give up and go home.

The argument reaches an impasse, and a conflict needs to be addressed to resolve it. There are two possible conclusions here, depending on how the conflict is resolved:

6. On the whole, I'd rather go home.

$6'$. On the whole, I'd rather go on hiking.

The main purpose of Steps 1–5 is to identify the conflict.

I'm not altogether happy with the theory presented in [Thomason, 2000], but I still believe that the practicalization of desires is an important part of practical reasoning that provides opportunities for using logic to good advantage.

2.5 Intention formation

The product of successful means-end deliberation will be an intention, taking the form of commitment to a plan. But the deliberation would not get started without a goal—and I see no difference between a goal and a provisional and (perhaps very general and sketchy) intention. Often, even in human agents, these goals come from habits, or from compliantly accepted instructions from other agents.

But sometimes goals arise internally, as outcomes of deliberation. The hiker in Section 2.4 provides an example. If the conclusion of the reasoning is a practicalized desire to turn back and head for home, commitment to the conclusion will produce an intention, which may even become a goal for means-end reasoning. ("How am I to get home?")

This is why practicalization can be an important component of practical reasoning, especially if the reasoner is an autonomous human being.

2.6 What to do now?

In the life of an autonomous agent, moments will arise when there is scope for new activities. These opportunities need to be recognized, and an appropriate task needs to be selected for immediate execution. A busy agent with many goals and a history of planning may have a ready-made agenda of tasks for such occasions; but even so, it may take reasoning to select a task that is rewarding and appropriate. I do not know if any useful work has been done on this reasoning problem.

2.7 Scheduling, execution and engagement

Some of the examples in Section 1.1 were intended to illustrate the point that there can be deliberation even in the execution of physically demanding, real-time tasks. And there can be such a thing as overplanning, since the plans that an agent makes and then performs will need to be adjusted to circumstances, and more detailed plans will require more elaborate adjustments.

Also, not all intentions are immediate. Those that are not immediate need to be invoked when the time and occasion are right.

There has been a great deal of useful work on these topic in AI; just one one recent example is [Fritz, 2009].

2.8 Framing a practical problem

Leonard Savage's "Small worlds problem" is replicated in the more qualitative setting of means-end deliberation. A means-end reasoning problem requires (at least) a set of actions, a description of the initial conditions, and a goal. But, even in complex cases, formulations of planning problems don't include every action an agent might perform, or every fact about the current state of the world. Somehow, a goal (like "getting to the airport") has to suggest a method of distinguishing the features of states (or "fluents") and the actions that are relevant and appropriate.

I'm sure that ontologies would be helpful in addressing this problem, but other than this I have very little to say about it at the moment.

BIBLIOGRAPHY

[Allen *et al.*, 1990] James Allen, James Hendler, and Austin Tate, editors. *Readings in Planning.* Morgan Kaufmann, San Mateo, California, 1990.

[Anscombe, 1958] G.E.M. Anscombe. *Intention.* Blackwell Publishers, Oxford, 1958.

[Asada *et al.*, 1999] Minoru Asada, Hiroaki Kitano, Itsuki Noda, and Manuela Veloso. RoboCup today and tomorrow—what we have learned. *Artificial Intelligence*, 110(2):193–214, 1999.

[Baier and McIlraith, 2008] Jorge A. Baier and Sheila A. McIlraith. Planning with preferences. *The AI Magazine*, 29(4):25–36, 2008.

[Bates, 1994] Joseph Bates. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994.

[Belnap, Jr. *et al.*, 2001] Nuel D. Belnap, Jr., Michael Perloff, and Ming Xu. *Facing the Future: Agents and Choices in Our Indeterminist World.* Oxford University Press, Oxford, 2001.

[Blythe, 1999] Jim Blythe. An overview of planning under uncertainty. In Michael Wooldridge and Manuela Veloso, editors, *Artificial Intelligence Today*, pages 85–110. Springer-Verlag, Berlin, 1999.

[Boutilier *et al.*, 2003] Craig Boutilier, Ronen I. Brafman, Carmel Domschlak, Holger H. Hoos, and David Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2003.

[Brafman *et al.*, 2006] Ronen I. Brafman, Carmel Domshlak, and Solomon E. Shimony. On graphical modeling of preference and importance. *Journal of Artificial Intelligence Research*, 25:389–424, 2006.

[Bratman *et al.*, 1988] Michael E. Bratman, David Israel, and Martha Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.

[Chellas, 1969] Brian Chellas. *The Logical Form of Imperatives.* Perry Lane Press, Stanford, California, 1969.

[Church, 1959] Alonzo Church. *Introduction to Mathematical Logic, Vol. 1.* Princeton University Press, Princeton, 1959.

[Domschlak, 2002] Carmel Domschlak. *Modeling and Reasoning about Inferences with CP-Nets.* Ph.d. dissertation, Ben-Gurion University of the Negev, Be'er Sheva, 2002.

[Doyle, 2004] Jon Doyle. Prospects for preferences. *Computational Intelligence*, 20(2):111–136, 2004.

[Fagin *et al.*, 1995] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge.* The MIT Press, Cambridge, Massachusetts, 1995.

[Fitting, 2009] Melvin Fitting. Intensional logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford, California, 2009.

[Fox and Clarke, 1991] J. Fox and M. Clarke. Towards a formalization of arguments in decision making. In *Proceedings of the 1991 AAAI Spring Symposium on Argument and Belief*, pages 92–99. AAAI, 1991.

[Fritz, 2009] Christian Fritz. *Monitoring the Generation and Execution of Optimal Plans*. Ph.d. dissertation, University of Toronto, Toronto, 2009.

[Gabbay and Woods, 2005] Dov M. Gabbay and John Woods. The practical turn in logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic, Volume XIII*, pages 1–122. Springer-Verlag, Berlin, 2 edition, 2005.

[Geach, 1963] Peter T. Geach. Imperative inference. *Analysis*, 23, Supplement 1(3):37–42, 1963.

[Georgeff *et al.*, 1999] Michael Georgeff, Barney Pell, Martha Pollack, Miland Tambe, and Michael Wooldridge. The belief-desire-intention model of agency. In Jörg P. Müller, Munidar P. Singh, and Anand S. Rao, editors, *Intelligent Agents V: Agents Theories, Architectures, and Languages*, pages 1–10. Springer-Verlag, Berlin, 1999.

[Goldsmith and Junker, 2008] Judy Goldsmith and Ulrich Junker. Preference handling for artificial intelligence. *The AI Magazine*, 29(4):9–12, 2008.

[Green, 1997] Mitchell Green. The logic of imperatives. In E. Craig, editor, *The Routledge Encyclopedia of Philosophy*, pages 717–21. Routledge, New York, 1997.

[Hansson, 2001] Sven Ove Hansson. Preference logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic, Volume IV*, pages 319–394. Kluwer Academic Publishers, Amsterdam, 2 edition, 2001.

[Harel *et al.*, 2000] David Harel, Dexter Kozen, and Jerzy Tiuryn. *Dynamic Logic*. The MIT Press, Cambridge, Massachusetts, 2000.

[Hofstadter and McKinsey, 1939] Albert Hofstadter and J.C.C. McKinsey. On the logic of imperatives. *Philosophy of Science*, 6:446–457, 1939.

[Jones and Carmo, 2002] Andrew J.I. Jones and José Carmo. Deontic logic and contrary-to-duties. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic, Volume VIII*, pages 265–344. Kluwer Academic Publishers, Amsterdam, 2 edition, 2002.

[Jørgensen, 1937-1938] Jørgen Jørgensen. Imperatives and logic. *Erkenntnis*, 7:288–296, 1937-1938.

[Kahneman and Tversky, 1979] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

[Keeney and Raiffa, 1976] Ralph H. Keeney and Howard Raiffa. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley and Sons, Inc., New York, 1976.

[Kenny, 1966] Anthony J. Kenny. Practical inference. *Analysis*, 26(3):65–75, 1966.

[Lemmon, 1965] Edward John Lemmon. Deontic logic and the logic of imperatives. *Logique et Analyse*, 8:39–71, 1965.

[Lewis, 1979] David K. Lewis. A problem about permission. In Esa Saarinen, Risto Hilpinen, Ilkka Niiniluoto, and Merrill Province Hintikka, editors, *Essays in Honour of Jaakko Hintikka*. D. Reidel Publishing Co., Dordrecht, Holland, 1979.

[Lifschitz *et al.*, 2000] Vladimir Lifschitz, Norman McCain, Emilio Remolina, and Armando Tacchella. Getting to the airport: The oldest planning problem in AI. In Jack Minker, editor, *Logic-Based Artificial Intelligence*, pages 147–165. Kluwer Academic Publishers, Dordrecht, 2000.

[McCarthy and Hayes, 1969] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In Bernard Meltzer and Donald Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, 1969.

[Millgram, 2001] Elijah Millgram. Practical reasoning: The current state of play. In Elijah Millgram, editor, *Varieties of Practical Reasoning*, pages 1–26. The MIT Press, Cambridge, Massachusetts, 2001.

[Minsky, 2006] Marvin Minsky. *The Emotion Machine*. Simon & Schuster, New York, 2006.

[Newell, 1992] Allen Newell. *Unified Theories of Cognition.* Harvard University Press, Cambridge, Massachusetts, 1992.

[Reiter, 2001] Raymond Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems.* The MIT Press, Cambridge, Massachusetts, 2001.

[Ros *et al.*, 2009] Raquel Ros, Josep Llus Arcos, Ramon Lopez de Mantaras, and Manuela M. Veloso. A case-based approach for coordinated action selection in robot soccer. *Artificial Intelligence*, 173(9–10):1014–1039, 2009.

[Ross, 1941] Alf Ross. Imperatives and logic. *Theoria*, 7:53–71, 1941. Reprinted with minor changes, in *Philosophy of Science*, Vol. 11 (1944), pp. 30–46.

[Russell and Wefald, 1991] Stuart J. Russell and Eric Wefald. *Do the Right Thing.* The MIT Press, Cambridge, Massachusetts, 1991.

[Sanner and Boutilier, 2009] Scott Sanner and Craig Boutilier. Practical solution techniques for first-order MDPs. *Artificial Intelligence*, 173(5–6):748–788, 2009.

[Savage, 1972] Leonard Savage. *The Foundations of Statistics.* Dover, New York, 2 edition, 1972.

[Shanahan, 1997a] Murray Shanahan. A logical formalisation of Ernie Davis' egg cracking problem. Unpublished manuscript, Imperial College London, 1997.

[Shanahan, 1997b] Murray Shanahan. *Solving the Frame Problem.* The MIT Press, Cambridge, Massachusetts, 1997.

[Simon and Schaeffer, 1992] Herbert A. Simon and Jonathan Schaeffer. The game of chess. In Robert J. Aumann and Sergiu Hart, editors, *Handbook of Game Theory with Economic Applications, Vol. 1*, pages 1–17. North-Holland, Amsterdam, 1992.

[Simon, 1947] Herbert A. Simon. *Administrative Behavior.* The Macmillan Company, New Yori, 1947.

[Skyrms, 1990] Brian Skyrms. *The Dynamics of Rational Deliberation.* Harvard University Press, Cambridge, Massachusetts, 1990.

[Solomon, 1976] Robert C. Solomon. *The Passions: The Myth and Nature of Human Emotion.* Anchor Press, New York, 1976.

[Thomason, 2000] Richmond H. Thomason. Desires and defaults: A framework for planning with inferred goals. In Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, editors, *KR2000: Principles of Knowledge Representation and Reasoning*, pages 702–713, San Francisco, 2000. Morgan Kaufmann.

[Thomason, 2002] Richmond H. Thomason. Preferences as conclusions. In Ulrich Junker, editor, *Preferences in AI and CP: Symbolic Approaches*, pages 94–98. AAAI Press, Menlo Park, California, 2002.

[Tversky and Kahneman, 1981] Amos Tversky and Daniel Kahneman. The framing of decisions and the psychology of choice. *Science, New Series*, 211(4481):453–458, 1981.

[Visser and Burkhard, 2007] Ubbo Visser and Hans-Dieter Burkhard. RoboCup: 10 years of achievements and future challenges. *The AI Magazine*, 28(2):115–132, 2007.

[von Wright, 1963] Georg Henrik von Wright. Practical inference. *The Philosophical Review*, 72:159–179, 1963.

[von Wright, 1983] Georg Henrik von Wright. *Practical Reason: Philosophical Papers, Volume 1.* Cornell University Press, Ithaca, 1983.

[Weirich, 2004] Paul Weirich. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances.* Oxford University Press, Oxford, 2004.

[Williams, 1963] B.A.O. Williams. Imperative inference. *Analysis*, 23, Supplementary 1(3):30–36, 1963.

[Wooldridge, 2000] Michael J. Wooldridge. *Reasoning about Rational Agents.* Cambridge University Press, Cambridge, England, 2000.

M. ABRAHAM, D. M. GABBAY AND U. SCHILD

PRINCIPLES OF TALMUDIC LOGIC

The topics addressed in this chapter deal with the logic of Halacha — Jewish law, and in particular with the logic of the Talmud. In this preface we intend to analyse the nature and characteristics of the Talmud, and determine its place within the Halachic context. Next, we shall attempt to define our research aims with respect to these topics, as expressed in the sections of this chapter.

1 A GENERAL SURVEY OF THE HALACHIC TRADITION: FROM THE CREATION OF THE WORLD UNTIL THE GIVING OF THE TORAH

According to Jewish tradition the origin of the Jewish people is in the ancient East, where Abraham the son of Terach was born and lived. Abraham was the first monotheist. According to the usual time-scale of Jewish tradition he was born in 1812 BCE. Abraham communicated with the Creator, but did not yet know the specifics of the Torah (the Pentateuch) to be given on Mount Sinai several hundred years after his death. Talmudic Midrashim (homiletic teachings on the Bible) state that Abraham fulfilled commands of the Torah, perhaps even all the commands. Formally, however, it is assumed that the Halacha, in its traditional significance as a set of norms that each Jew must obey, did not exist yet in his time.

Nevertheless, first concepts of Halacha existed even before that. The Halachic era began two thousand years earlier, when Adam was given six commandments, and a little later when Noah was given a seventh commandment. These are seven universal obligations for all human beings, whether of the Jewish faith or not, and they are denoted in the Halachic jargon: "The seven commandments of Noah's sons".

The beginning of Jewish Halacha is with Abraham, who received the command of circumcision. After him, the process continues with the Egyptian Diaspora, where Amram, the father of Moses, receives some further early commandments. The process continues at Marah, during the wanderings of the Jews in the desert towards the land of Israel. At Marah the Israelites received an additional three commandments from God. (There are different traditions identifying these three commandments). The giving of the Torah by God takes place at Mount Sinai, where Moses receives the Torah from God and passes it to his disciples and on to the Children of Israel in all generations.

D. M. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*,
Handbook of Philosophical Logic 18, https://doi.org/10.1007/978-3-319-97755-3_4

This is how Maimonides[1] describes it (The Book of Judges, Kings and Wars, chapter IX, 1):[2]

> Six precepts were given to Adam: prohibition of idolatry, of blasphemy, of murder, of robbery, and the command to establish courts of justice. Although there is a tradition to this effect-a tradition dating back to Moses, our teacher, and human reason approves of those precepts-it is evident from the general tenor of the Scriptures that he (Adam) was bidden to observe these commandments./ An additional commandment was give to |Noah: prohibition of (eating) a limb from a living animal, as it is said: *Only flesh with the life thereof, which is the blood thereof, shall ye not eat* (Gen. 9:4). Thus we have seven commandments. So it was until Abraham appeared who, in addition to the aforementioned commandments, was charged to practice circumcision. Moreover, Abraham instituted the Morning Service, Isaac set apart tithes and instituted the Afternoon Service, Jacob added to the preceding law (prohibiting) the sinew that shrank, and inaugurated the Evening Service. In Egypt Amram was charged to observe other precepts, until Moses came and the Law was completed through him.

Moses led the Children of Israel out of Egypt, and fifty days after the Exodus the people arrived at Mount Sinai and received the Torah from Moses.

It is important to understand that despite the above description of stepwise receipt of the Halacha, the customary Halachic assumption is that the Sinaitic revelation gives the validity and creates the Halachic obligation. This is what Maimonides writes (The Book of Judges, Kings and Wars, chapter VIII, 11):[3]

> A heathen who accepts the seven commandments and observes them scrupulously is a “righteous heathen”, and will have a portion in the world to come, provided he accepts them and performs them because the Holy One, blessed be He, commanded them in the Law and made known through Moses, our teacher, that the observance thereof had been enjoined upon the descendants of Noah even before the Law was given. But if his observance thereof is based upon a reasoned conclusion he is not

[1]Moses ben-Maimon, called Maimonides and also known as Rambam (Hebrew acronym for “Rabbi Moshe ben Maimon”), was a preeminent medieval Jewish philosopher and one of the greatest Torah scholars and physicians of the Middle Ages. He was born in Córdoba, Spain, 1135, and died in Egypt (or Tiberias) on December 12, 1204. He was a rabbi, physician and philosopher in Morocco and Egypt.

[2]The Code of Maimonides, Book Fourteen, The Book of Judges, Yale Judaica Series, Yale University Press, 1949

[3]As above.

> deemed a resident alien, or one of the pious of the Gentiles, but one of their wise men.

Maimonides states that even a Ger Toshav (a resident alien), i.e. an alien who observes the seven commandments of Noah's sons, must do that as part of the Sinaitic tradition. Observance of commandments through logical acceptance of their importance and truth is not of religious merit, but only of moral value. The commentators at that place explain that this principle also holds for Jews, with respect to the observance of commandments in general.

Maimonides (Commentaries on the Mishna[4], Chulin 7:6) states what at first sight seems to be a similar principle':[5]

> You must know that whatever we do or refrain from doing today, we do only because of God's command by way of Moshe, and not because of God's command to the prophets that preceded him. For example, we refrain from eating a limb removed from a living animal not because God forbade this to the descendants of Noah, but because God forbade it to us when He commanded us at Sinai that a limb removed from a living animal continues to be forbidden... You see that [the Sages] said (Makkot 23b): "Six hundred and thirteen mitzvot were told to Moshe at Sinai," and all these are included among the mitzvot.

Here Maimonides states that the commandments including the ones from before the Sinaitic revelation should also be obeyed because of the renewed obligation at that point. Thus Maimonides seems to repeat what he said in the Book of Judges, Kings and Wars quoted above, but a closer observation shows that these are really two different statements.

In order to understand this we must mention a concept coined by Hans Kelsen, a jurist and legal philosopher: Positivism. Kelsen, the positivist, saw the legal system as a logical system of hierarchical norms. Each norm depends on a norm higher in the hierarchy. At the top of the normative pyramid stands the "basic norm" which gives validity to the entire system. In the context of national laws the basic norm can be the obligation to obey the legislating body, or carry out the wishes of the electors as expressed by the legislation, etc. In the Halachic context the basic norm may be taken as the Halachic obligation that Jews took upon themselves at Mount Sinai. The Sages expressed it as follows (Tractate Nedarim 8a and similar places[6]): "As if we all swore at Mount Sinai to obey the Torah".

---

[4]The Mishna or Mishnah (Hebrew: "repetition", from the verb shanah, "to study and review", also "secondary") is the first major written redaction of the Jewish oral traditions called the "Oral Torah"

[5]Free translation.

[6]where nothing else is stated the reference is to the Babylonian Talmud.

Talmud Bavli, The Schottenstein Edition, Mesorah Publications, Ltd, NY, 2008

The words of Maimonides in the Commentaries on the Mishna[7] quoted above state that that the obligation to obey the Halacha is founded on the basic norm: What was given by God at Mount Sinai (and not earlier or later). So what about a person who obeys the law, but not because of the basic norm? For example, assume a citizen of Israel crosses an intersection only at a green light and pays his taxes lawfully, but he does this not because he accepts the legislation of the Knesset (parliament), but because he considers the norms as morally correct. Is there a legal defect in his behaviour? Certainly not. Kelsen's basic norm does not appeal to the citizens but to the government; the basis of a prosecution of a citizen who does not obey the law is the basic law. This is the justification for punishing him. He must obey the law, and it does not matter what his motivation may be.

Maimonides in the book of Judges, Kings and Wars says something else. He states that if somebody obeys the obligations owing to reason and logic and not because of his obligation to the basic norm (The Torah at Mount Sinai), then his actions are not considered as Mitzvot (fulfillment of commandments). A person who puts on Tefillin (phylacteries) without believing in God, or without believing and feeling obliged to Mount Sinai, is not fulfilling a commandment. In principle he must put on the Tefillin a second time.

We see that the basic norm is central to Halacha and more important than in other legal systems. Here there is an appeal to the individual, and not just a theoretical justification for the government to act against the wrongdoer. In Halacha an action according to the basic norm is necessary for the action to be considered a Mitzva.

1.1 From Mount Sinai and on: The Development of the Oral Code

The Jewish tradition states that at Mount Sinai Moses received two parts of the Torah: The Written Law (The five books of Moses — the Pentateuch) and additional principles and commentaries called the Oral Law.

From that point and on there is a chain of transmission from a teacher to his disciple, as described in the Mishna (Tractate Avot, chapter I, 1-2):[8]

> Moses received the Torah from Sinai, and transmitted it to Joshua, and Joshuah to the Elders, and the Elders to the Prophets, and the Prophets transmitted it to the Men of the Great Assembly. . . : Shimon HaTzaddik was [one] of the remnants of

[7]The Mishna or Mishna is the first major written redaction of the Jewish oral traditions called the "Oral Torah".

[8]The Mishna, Artscroll Mishna Series, Seder Nezikin, Vol. IV, Avos, Mesorah Publications, Ltd, NY, 2007.

> the Great Assembly... Antigenos, leader of Socho, received [the Mesorah] from Shimon HaTzaddik...: Yose ben Yoezer [the] leader of Tzredah and Yose ben Yochanan [the] leader of Jerusalem received [the Mesorah] from them... : Yehoshua be Perachyah and Nittai of Arbel received [the Mesorah] from them...: Yehudah be Tabbai and Shimon ben Shatch received [the Mesorah] from them...: Shemayah and Avtalion received [the Mesorah] from them...: Hillel and Shammai received [the Mesorah] from them...: Rabbi Yochanan ben Zakkai received [the tradition] from Hillel and Shammai...: Rabbi Yochanan be Zakkai had five disciples. They were: R'Eliezer ben Hyrkanos, R'Yehoshua be Chananiah, R'Yose the Kohen, R'Shimon ben Nesanel, and R'Elazar ben Arach.

This description is abridged, and in parallel sources one can find more detailed descriptions of the intermediate steps. At the end of the Tannaic[9] period it is decided to write down the principles of the Oral Law, and thus the Mishna is written by Rabbi Yehuda HaNassi (at the beginning of the third century CE).

After the Tannaic period outlined in the Mishna above, the tradition continues as described in the literature of the Rishonim[10], prefaces by Maimonides, and others: The Amoraim[11] created two works of Talmud: The Babylonian Talmud (often called the Gemara) and the Jerusalem (= Palestinian, = Israeli) Talmud. According to tradition the Jerusalem Talmud was completed by Rabbi Yochanan at the end of the fourth century CE, and the Babylonian Talmud was completed by Rav Ashi and Ravina at beginning of the sixth century CE. It is usually assumed that the finalising and editing of the two Talmuds was done over a period of several hundred years after the date given by tradition. After the Amoraim came the Savoraim[12], then the Gaonim[13] and then the Rishonim[8] (= sages of the Middle Ages). During the period of the Rishonim there was an large amount of Halachic activity, mainly interpretation of the Scriptures and the Babylonian Talmud. Exceptional during this period is Rabbi Moshe ben Maimon,

[9]Tannaim (plural of Aramaic tanna,=one who studies or teaches), Jewish sages of the period from Hillel to the compilation of the Mishna .They functioned as both scholars and teachers, educating those in the synagogues as well as in the academies. Their opinions are found either in the Mishna or as collected in the Tosefta .

[10]Rishonim, leading rabbis who were deciders of Jewish law and lived between 1050 and 1500 CE

[11]Amoraim, renowned Jewish scholars who "said" or "told over" the teachings of the Oral law, from about 200 to 500 CE in Babylonia and the Land of Israel

[12]Savoraim (s. Savora, Aramaic "a reasoner") are the leading rabbis living from the end of period of the Amoraim (around 500 CE) to the beginning of the Geonim (around 700 CE).

[13]Gaonim (also transliterated Geonim) were the presidents of the two great Talmudic Academies of Sura and Pumbedita, in Babylonia, and were the generally accepted spiritual leaders of the Jewish community world wide in the early medieval era

Maimonides[1]. He created a monumental work, the Mishne Torah (the Code of Maimonides). In this book of fourteen parts Maimonides codifies all Halachic writings from the Talmudic and extra-Talmudic sources and the period of the Gaonim. This is a unique work in the long history of the Halacha. It is the only work that contains the entire Halachic corpus in a unified, methodical and organised structure. Other works do not cover all areas of the Halacha, and certainly do not classify the material and settle disputes in a systematic manner.

The beginning of Modern Times is the beginning of the Acharonim, leading rabbis and Poskim (Jewish legal decisors) living from roughly the 16th century to the present. During the period ending the Middle Ages and the beginning of Modern Times a major Halachic work, the Shulchan Aruch (Set Table), was created by Yosef Karo[14] in Safed and by Rabbi Moshe Isserles[15] in Krakow. This work which summarises the major commandments relating to everyday life at the time, does not include commandments that deal with the service in the (no more existing) Temple, with questions of cleanliness and purification (which are not observed today). Halachic developments continue of course also today.

1.2 The Work of the Sages: Types of Halacha

The Halacha was given by God at Mount Sinai, but it expands and is elaborated all the time, to this very day. The Halacha is a legal system, and as such it is supposed to supply answers to the legal needs of Jewish society. It follows that each link in the chain of transmission described above is responsible for several tasks: Transmitting the knowledge accumulated so far, making new regulation and rabbinical decrees, responding to questions by ordinary people and passing judgments in rabbinical courts. However, it is true that the Halacha operated and developed at places and times when it has no full autonomy, under the rule of different legal systems.

It is customary to divide Halacha into two major categories: Halacha mainly revealed at Mount Sinai (Halacha DeOraita) and Rabbinical Halacha. It is a common mistake to believe that this distinction is entirely chronological. Halacha created by Biblical hermeneutics (to be dealt with below), or interpretation of oral traditions given to Moses at Mount Sinai is Halacha DeOraita. Halacha created by rabbinical legislation (and not by Biblical exegesis) is Rabbinical Halacha.

This means that Halachic items are created all through history, join the Halachic corpus and their status may be like oral traditions given to Moses at Mount Sinai., or their status is as rabbinical legislation.

[14]Joseph ben Ephraim Karo (also spelled Yosef Caro, or Qaro), Toledo, Spain 1488 – Safed, Israel 1575

[15]Moses Isserles (also spelled Moshe Isserlis, called the Remah), Kraków, Poland, 1520–May 11, 1572

Apart from the theoretical difference there is also a Halachic distinction between the two kinds of Halacha. Halacha DeOraita is more stringent than Rabbinical Halacha, and this difference has Halachic implications. Consider for example, the Halacha of Doubt. If Doubt concerning a Halacha DeOraita is strict, a person must exceed the bare requirements of the Halacha: If a person is about to eat meat, and he is not sure whether the meat is pork or veal, he must not eat it. In the case of Rabbinical Halacha one is more lenient.

1.3 The Anarchistic Character of Halacha

Based on our description above one would expect the Halacha to become well organised over time, and that fixed and constant Halachic procedures would develop. Surprisingly, Halacha has an anarchistic element that has not been eliminated through history. It has even been asserted that a modern legal system cannot be based on the Halacha because of this anarchistic trait[16].

Many consider this fundamental trait to be part of Jewish identity and nature, which is original, rebellious and argumentative, which is divisive into different lines of thought, and does not accept authority. Judaism in general and Halacha in particular are based on negotiation more than on a closed set of principles and obligatory specifics.

The anarchistic trait of the Halacha is expressed in several ways. It is found in the dispute among Halachic experts, but even more so it can be seen in the nature of the canonical works and their stature. We have already mentioned that throughout history we do not find any attempts to edit the various parts of Halacha and organise them in a classified manner. The attempt by Maimonides is exceptional, and his oeuvre was not acknowledged as binding. His work is considered one of the most important Halachic sources, but it is not a compulsory canonical codex.

If one wishes to speak about a binding canonical codex it must be the Talmud (especially the Babylonian Talmud). But many questions arise when the nature of this text is considered. It is not a codex in any sense. It is a collection of fragments (sugiot) that clarify various commandments. Among them are aggadatic (homiletic) fragments, stories, moral lessons, etc. Also the purely Halachic parts are formulated as discussion and debate among the Sages. They cite different sources, and only seldom do they state a definitive Halacha. In most cases the fragment ends by simply stating the different opinions and their Halachic conclusions. It is a description of discussions at various places, about different topics in different formulations. Sometimes there are contradictions between different Talmudic fragments,

[16] Hanina Ben-Menahem. Judicial Deviation in Talmudic Law: Governed by Men, Not by Rules. Jewish Law in Context, vol. 1. New York: Harwood Academic Publishers, 1991. xi, 220 pp.

and also the textual versions are problematic. The Talmud is open textured, not what one may expect from a canonical text that aspires to be a codex.

So the central Halachic canon is only a collection of discussions and Halachic arguments, which also include elements that only relate to the discussions in an associative manner. It therefore looks very strange that such a text should be accepted as the central binding canon of a normative legal system. What is gained by this canonisation? What does it mean? As there is nothing final, as the questions remain open, what is canonical about this canon? In order to answer that question we must first consider the historical background leading to the acceptance of the Talmud as a binding canonical text.

The history of the Halacha as schematically described above can be considered as a chain of links; each link is a (not exactly defined) time period, with a name: Prophets, Men of the Great Assembly, the Couples, Tannaim, Amoraim, Savoraim, Gaonim, Rishonim and Acharonim. In some cases there is a Halachic significance in the passing from one period to the other: Amoraim do not disagree with Tannaim. A Tannaic source that contradicts an Amorai is usually considered a definitive proof against the Amorai. Similarly post-Talmudic sages do not disagree with the Talmud, most of the Rishonim do not usually disagree with the Gaonim (though the feature is less apparent), and similarly with Acharonim and Rishonim.

It would seem that this picture reduces the anarchistic nature we described above. However, we must remember that in each period there were several sages, each with his opinions, and they created different sources. Therefore, the general obligation towards a certain period does not mean much from a practical aspect. Nevertheless, the question remains what is the nature of the prohibition of sages from one period to disagree with sages from a previous period. This is particularly important in the (not too many) cases where the Talmudic fragment ends with a definite conclusion. This conclusion is binding on the following generations.

Maimonides (The Book of Judges, Rebels, chapter II, 1) states the following:

> If the Great Sanhedrin, by employing one of the hermeneutical principles, deduced a ruling which in its judgment was in consonance with the Law and rendered a decision to that effect, and a later Supreme Court finds a reason for setting aside the ruling, it may do so and act in accordance with its own opinion, as it is said: *and unto the judge that shall be in those days* (Deut. 17:9), that is, we are bound to follow the directions of the court of our own generation.

It would seem that there is no restriction on the sages of one generation to formulate the Halacha as they wish. Their capability of disagreeing with their predecessors does not require that they are greater than the latter in learning, or any other requirement[17].

Rabbi Yosef Karo, the author of the Shulchan Aruch, in his commentary on Maimonides (Kesef Mishne[18]: On The Book of Judges, Rebels, chapter II, 1) considers the question why Amoraim do not disagree with the Tannaim, and why the Gaonim and Rishonim do not disagree with the sages of the Talmud: If there are no limitations on disagreement with previous generations, it is not clear why this does not happen in practice.

One might expect the explanation lies in praising the greatness of early generations. Thus we find several times in the Talmud, i.e. (Shabbat 112b):

> R' Zeira said in the name of Rava bar Zimona: If the early ones were sons of angels, we are sons of men; and if the early ones were sons of men, we are like donkeys.

but surprisingly, Kesef Mishne (loc. cit.) chooses another explanation. Its argument is that the authority of the Talmud is because we have decided not to challenge it. Thus, it is a technical acceptance only. In principle it could be possible to contradict any sage in any generation, but at the end of the Talmudic period the sages decided to accept the Talmud as a canonical corpus, not to be contradicted.

If there is no real Halachic constraint, why did the sages of that generation decide to stray from the Halachic anarchism and establish an obligatory text? The answer lies in processes that took place towards the end of the first Millennium CE (the beginning of the Middle Ages).Until then, Jews had been living in the Babylonian Diaspora for about a thousand years, they had legal autonomy, and an organised hierarchy headed by the Exilarch[19], who functioned in place of a king or president. At that time, this structure began to crumble. Jews began to spread all over the world, and form small communities at various distant places. There was therefore a danger that the Halacha would lose its coherence. Every little community, which may not have a person qualified in Jewish law, and no control over the ways of interpreting the law, could create some undisciplined legal interpretations. The Halacha would disintegrate, and Jewish social and legal cohesiveness would be destroyed. Therefore it was decided at that time to establish a framework for the further development of the Halacha. This framework is the Talmud.

It is important to understand that while the sages of those generations found it acceptable to deviate from the usual custom and establish a bind-

[17] In the next paragraph Maimonides introduces limitations on the capability of disagreeing with previous religious courts, but this relates only to Rabbinical legislation and regulations, and not to Halacha DeOraita, i.e. Biblical exegesis.

[18] Kesef Mishne, first printed Venice, 1574

[19] Exilarch (Hebrew: Rosh Galut, lit. "head of the exile") refers to the leaders of the Diaspora Jewish community following the deportation of the population of Judah into Babylonian exile after the destruction of the kingdom of Judah.

ing canon, it seems that they also wished to preserve the open nature of the Halacha. The decision was therefore to select an open work like the Talmud as a binding canon. This step established a framework for discussion and development of the Halacha, enabled the discourse among communities and among sages all over the world, as indeed has been observed in all following generations. At the same time there is also room for considerable interpretational freedom. The result of this decision is the flowering of different communities and different interpretations of the sources of Halacha. During the thousand years since then the sages have acted autonomously, without a central Halachic authority, and without authority of enforcement. At the same time they have managed to keep up a debate among themselves. We believe this is a unique phenomenon, which has no analogue in the history of human civilisation.

Part of the debate is about the method of discussion itself. There are many conflicts about how to deal with controversy and different Halachic opinions. The debate is open, while at the same time its coherence is kept within the framework of the Talmud.

In order to illustrate the tension between the need for coherence and the wish to maintain freedom of opinions and dialectic, we shall describe a critical moment in the progress of Halachic disputation, which is found in several fragments of the Talmud and parallel writings.

1.4 The Way of Settling Controversy

Tractate Berachot 28a describes the forced abdication of R' Gamliel from the presidency[20] and the appointment of R' Elazar ben Azaryah in his stead, and states:

> And any place wherein 'on that day' is used, it is a reference to that day R' Elazar ben Azaryah was installed as Nasi

Let us attempt to clarify the importance of that day, which the Talmud has singled out in this manner.

In the second generation of sages in Yavneh a revolution took place, which has great importance for the further development of the Oral Law. The event is mentioned in several places in the Talmud, usually in a very dramatic way[21], but the historical correspondence between the descriptions is not clear at first sight. The words of Jewish Law are sparse in one place and rich in other places, so let us begin by examining a strange phenomenon in Tractate Avot.

[20]The Sanhedrin (Greek: synedrion, "sitting together," hence "assembly" or "council") was an assembly of twenty-three judges appointed in every city in the Biblical Land of Israel. The Great Sanhedrin was the supreme court of ancient Israel made of 71 members. The Nasi was the president of the Sanhedrin including when it sat as a criminal court.

[21]Tractate Baba Metziah 59, Tractate Berachot 28, Tractate Sanhedrin 68 and 101 and Tractate Chagiga 3.

As we saw above, the first chapter of Tractate Avot describes the transmission of the Torah from Moses to Yehoshua, to the Elders, etc. This process ends with the fifth couple: Hillel and Shammai (see quote on page 6). After that the Mishna brings sayings of sages from several generations, until chapter 2:9. There the description of the transmission is taken up again with the words: "R' Yochanan ben Zakkai received [the transmission] from Hillel and Shammai".

Immediately after that the description of the transmissions ends. In the next Mishna it says: "Rabbi Yochanan be Zakkai had five disciples. They were: R'Eliezer ben Hyrkanos, R'Yehoshua be Chananiah, R'Yose the Kohen, R'Shimon ben Nesanel, and R'Elazar ben Arach", and this is the end of the narrative.

Until R' Yochanan ben Zakkai the process is described using the word 'received': "Moses received from Yehoshua and transmitted..." R' Yochanan ben Zakkai is still described as 'receiving'. After that the words 'receive' and transmit' are not associated anymore in the Tractate Avot with respect to the process of transmission. The description of the process does not include these words. It should be noted that in the generation after R' Yochanan ben Zakkai there is no single prominent person or couple as before. The process becomes crowded, each sage has several disciples.

R' Yochanan ben Zakkai is considered the first generation of Tannaim. It is well-known that he asked the Romans for Yavneh[22] and its sages. The next generation was the generation of R' Gamliel from Yavneh, who was the president, R' Eliezer ben Hyrkanos (R' Eliezer the Great'), brother-in-law of R' Gamliel, R' Yehoshua ben Chananyah, his friend and opponent, R' Elazar ben Azaryah, who was younger, and R' Akiva ben Yosef, who was older but still a disciple, first of R' Eliezer the Great and at then of R' Yehoshua.

The change from the times of the terminology of 'receipt' of a disciple from his teacher to the era of multiple disciples learning from a teacher indicates a significant process that the Oral Tradition goes through in the first generation in Yavneh (second generation of Tannaim).

The Babylonian Talmud describes a dramatic incident, where R' Yehoshua and R' Eliezer the Great disagree about the law concerning an oven of Achnai[23] (Tractate Baba Metziah 59b):

> On that day R' Eliezer advanced all the arguments in the world, but [the Sages] did not accept his arguments. [R' Eliezer] said to them: If the Halacha accords with me, let this Carob tree prove it... Let the water canal prove it... Let the walls of the

[22] After the destruction of the Second Temple in 70 CE, Rabban Yochanan Ben Zakkai moved the Sanhedrin[20] to Yavne. The Sanhedrin left Yavne for Usha in 80 CE and returned in 116 CE.

[23] The oven of Achnai is assembled rings of earthenware with sand in between them.

> study hall prove it ... whereupon a Heavenly echo went forth and proclaimed: What argument do you have with R' Eliezer, whom the Halacha follows in all places. R' Yehoshua stood on his feet and declared: It [The Torah] is not in Heaven ... We pay no heed to a Heavenly echo in matters of Halacha, because You already wrote in the Torah at Mount Sinai: According to the majority [the matter] shall be decided.

Let us first remark that the expression 'bo bayom' ('on that day', see above, page 12) appears here. The meaning is probably not just the day of the discussion, but the day when R' Gamliel was forced from the presidency. Further evidence for this can be brought from Tractate Berachot, which states that the entire Tractate of Eduyot was learned 'on that day'. Tractate Eduyot, chapter VII, 7 states:

> They testified about an [earthware] oven, that somebody cut into [horizontal] sections, and put sand between one section and the other, that it is susceptible to become unclean (tamei), because R'Eliezer rules such an oven clean (tahor).

In other words, the question of Achnai's oven was settled on the same day R' Gamliel was deposed from the presidency.

In order to understand the drama that took place that day, we must note that R' Eliezer the Great consistently represent the school that says that the Torah is all tradition, i.e., the law should be determined only by the information reaching us from the times of Moshe at Sinai. We shall show an example of this. In Tractate Sukkah 28a R' Eliezer states that he never said anything he had not heard from his teacher. R' Yochanan ben Zakkai praises R' Eliezer (Avot 2,8):

> Rabbi Eliezer the son of Hyrkanos is a cemented cistern that loses not a drop.

R' Eliezer himself said about himself: "Were all the seas ink, all the reed pens and all the men scribes they could not write all that I have studied" (Avot deRabbi Nathan, 80, 25). See also Tractate Sanhedrin 67-68 and 101, and many more examples. R' Eliezer was a great compiler of the knowledge of his teachers, and everything he said was in their name. His approach was one of tradition, he was a 'receiver'.

In the fragment in Tractate Baba Meziah 59b (Achnai's oven) R' Eliezer brings arguments that seem irrelevant: He makes miracles through the carob, the water canal, the heavenly voice. He attempts to prove that he is an expert, and therefore his viewpoint should be accepted. He does not prove the assertion itself. He does not bring reasons why his opinion is right, but reasons why he is a great man. This corresponds to his principled approach; with R' Eliezer the Halacha is decided because of his trustworthiness as conveyor of what he has learned from his teachers.

But R' Yehoshua, his friend and opponent, disagrees precisely on this point. He believes that the Halacha should be decided according to logic and wisdom: "[The Torah] is not in Heaven" (Deut 30:12). If there is no decision according to intellectual conviction there should be a vote, and the majority opinion should be the decision, as it is said; "yield to the majority" (Exodus 23:2). He does not accept the tradition that R' Eliezer presents without any reason and logic.

The end of the fragment (Tractate Baba Meziah 59a) describes how God himself says "My sons have vanquished me". Thus, according to the Talmudic tradition R' Yehoshua was victorious and overcame R' Eliezer. Even the Heavenly Voice did not help R' Eliezer, and the sages logically rejected his words. One may conclude that a 'Law of Debate' displaced a 'Law of Tradition', which had ruled until that day.

The President, R' Gamliel from Yavneh — friend and brother-in-law of R' Eliezer the Great — seems to have agreed to his concept of a 'Torah of Tradition'. This is why he, like R' Eliezer, was very careful to examine bearers of tradition. The Gemara tells us how he put guards at the entrance to the study hall (the academy), in order to allow entrance only to persons of like minds. According to R' Gamliel one has to ensure that the Torah is transmitted to persons who are trusted to transmit it on. When R' Gamliel was deposed, R' Elazar ben Azaryah was appointed in his place. He represented an opinion similar to that of R' Yehoshua: When the Torah is examined during a debate, it is not important to screen the participants according to their disposition and personality. Concepts are to be examined according to their nature, and not according to who expounds them. For that reason R' Elazar ben Azaryah decided to open up the study hall, and that day three hundred seats were added. The law of R' Elazar ben Azaryah is more democratic, as he does not examine the character of the disciples. According to him — and this was accepted from that moment and on — the Halacha is decided by debate and decisions are logical. No weight is given to the holiness and personality of the person expressing the opinion.

Apart from the discussion about the oven, the greater significance was the change from a 'Law of Tradition' to a 'Law of Debate'. This was a veritable revolution in the understanding of the Oral Code. A discussion about principles cannot be decided when the ruling system is 'Law of Tradition' — in a debate each side will be faithful to what he received from is teacher — and no conclusion can be reached.

In order to understand the timing and significance of this revolution one must consider the historical background. Maimonides in his preface to the Mishna describes how conflicts arose, when the disciples of Hillel and Shammai did not lend sufficient support to their teachers, tradition was lost and controversy arose (Tractate Sanhedrin 88b).

There had of course been disagreements before. The first one was in the Hellenistic period between Yose ben Yoezer and Yose ben Yochanan concerning the laying of hands on head of sacrifices on feast days. But in the times of Hillel and Shammai two schools of thought were created for the first time: Beit Hillel and Beit Shammai.

Such a situation cannot be resolved by the 'Law of Tradition' approach. One cannot decide between two schools based on different traditions. At that time the situation looked hopeless. It seemed like the Torah was about to crumble, and be lost to the world as a single and unique expression of Gods will. This is perhaps the way one should consider the account by the sages (Jerusalem Talmud, Tractate Shabbat, 81) saying that the disciples of Shammai actually murdered disciples of Hillel.

Hillel and Shammai belonged to the generation before R' Yochanan ben Zakkai. In Tractate Avot chapter I, 1-2 we saw that he received from both. In Tractate Sukkah he is described as the youngest of the disciples of the old Hillel. His disciple, R' Eliezer was already known as a Shammaite. So in the first generation in Yavneh a full scale controversy was already taking place. There was a danger of a general disintegration of the Torah.

The first generation of sages in Yavneh headed by R' Yehoshua ben Chananyah understood that such a situation warrants a real revolution in the approach to the Oral Tradition. It was necessary to develop a new approach in order to make decisions between the two schools that had arisen. In the case of open questions it was necessary to legitimise debate and decisions reached rationally or by plurality. This revolution, as described above in the case of Achnai's oven, was led by R' Yehoshua, and he was joined by his friend/disciple R' Elazar ben Azaryah, who as very young was appointed president instead of R' Gamliel.

The discussion in the case of Achnai's oven illustrates the type of issues in Tractate Eduyot, which were discussed 'on that day'. Right after his appointment R' Elazar be Azaryah brought all the open questions that could not be decided by the 'Law by Tradition' approach to a decision by debate and voting. Tractate Eduyot is somewhat exceptional in the Talmud as it does not have a definite subject which was learned 'on that day'. The tractate has, however, a very central theme, the new Oral Tradition and the decision of the issues that could not be decided before.

The Talmud describes how R' Eliezer the Great was ostracised by his friends/ disciples. Tractate Baba Metzia gives a heartbreaking description of his banishment, R' Akiva, his disciple, who volunteered to convey the bitter decision says to him: "My Teacher, it seems to me that your colleagues are removed from you" (Tractate Baba Metziah 59b), and they both wept. As we learn from Tractate Sanhedrin 88b, R' Eliezer stayed in isolation until his death. He stayed in Lud complaining that nobody pays him a visit, in order to learn from the vast amounts of Oral Tradition he knows.

It is not clear from the fragment itself what the banishment meant. It is not clear what sin R' Eliezer committed by daring to express a different Halachic viewpoint concerning the cleanliness of the oven. It is obvious that his demotion symbolises the end of the legitimacy of the Halachic approach that he represented: The "Law by Tradition". In view of the critical situation in the relationship among the sages (as described above), drastic measures were needed in order to introduce and settle the new face of the Oral Tradition in the study hall.

Also R' Gamliel, the brother-in-law of R' Eliezer, who held the same opinions, was deposed from the presidency in an unprecedented step. One may think the reason was the way R' Gamliel shamed R' Yehoshua (see e.g. Tractate Berachot 27b). However, that incident in reality shows the way R' Gamliel wished to impose the hierarchical approach as part of the 'Law by Tradition'. R' Yehoshua a rebel who went according to the logic and not the authority, gained the upper hand in the end. The Halachic anarchism was persecuted, but was victorious. The rebellion by R' Yehoshua in parallel to the acceptance by R' Gamliel of his authority (who was forced to desecrate what according to his view was the date of the Day of Atonement), may be denoted a 'Holy Rebellion'. R' Yehoshua was not interested in breaking totally with the past, but tried to convince his companions of the way of persuasion. In the words of the Mishna in Tractate Avot, chapter V, 17: "A debate for the sake of heaven will endure; but a debate not for the sake of heaven will not endure". The Mishna explains that disagreement and pluralistic views are good and important.

R' Gamliel accepted the rules of the game and was reinstalled. He rotated as president with R' Elazar ben Azaryah. R' Eliezer, on the other hand, staid in isolation until his death. He was not willing to change his approach of 'Law of Tradition'.

The development is illustrated by a fragment in Tractate Chagiga, 3a:

> There was once an incident involving R' Yochanan ben Broka and R' Eliezer (ben) Chisma, who went to visit R' Yehoshua in Pekiín. [R' Yehoshua] said to them: What novel teaching was expounded in the study hall today? They said to him: We are your disciples and we drink your waters. [R' Yehoshua] said to them: Even so, it is impossible for the scolars of the study hall without expounding a new teaching. Whose week was it to lecture in the study hall? It was the week of R' Elazar ben Azaryah. And on what subject was his discourse today?

R'Yehoshua says that 'Law of Debate" is a living and developing thing. It is not possible that there are no new developments in the study hall. So, while they wish to listen to his teachings, he wishes to learn from them.

The fragment in Tractate Chagiga, 3a continues with sayings by R' Elazar ben Azaryah., and R' Yehoshua continues (Tractate Chagiga, 3b):

> And he also started expounding "The words of the wise are like goads, and like nails well planted [are the sayings] of the masters of assemblies given from one shepherd"... Just as this plant is fruitful and multiplies, so the words of Torah cause one to be fruitful and multiply.
>
> The masters of assemblies – these are the wise scholars who sit in various groups and occupy themselves with Torah. There are those scholars who declare a thing ritually contaminated, and there are those who pronounce it clean. Those who prohibit and those who permit. Those who disqualify and those who declare fit. Perhaps a man will say: How can I ever learn Torah? Scripture states: All are given from one shepherd, one God gave them. One leader proclaimed them from the mouth of the Master, blessed is He. As is written: "And God spoke all these words". You make your ear like a mill-hopper, and acquire for yourself a discerning heart to hear intelligently the words of those that declare impure, and the words of those who declare pure, the words of those who prohibit, and the words of those who permit, and the words of those who disqualify and the words of those who declare fit.

Here we find the entire program of the Yavneh revolution, as carried out by R' Yehoshua and R' Elazar ben Azaryah. Next, R' Yehoshua stresses the importance for the Torah and the People of Israel (Tractate Chagiga, 3b):

> [R' Yehosua] then said in this language: It is not an orphaned generation that R' Elazar ben Azaryah dwells in.

The same fragment describes a meeting between R' Yose ben Durmaskis with his teacher R' Eliezer the Great (who sits excommunicated in Lud:

> There was once an incident involving R' Yose ben Durmaskis, who went to visit R' Eliezer in Lud. He said: What novel teaching was expounded in the study hall today? [R' Yose] told him: [The Sages] voted and decided Ammon and Moav must give the tithe of the poor in the seventh year. [R' Eliezer] replied to him: Yose stretch out your hands and darken your eyes. He stretched out his hands and darkened his eyes. R' Eliezer wept and declared: The secret of Hashem is to those who fear him and his covenant to inform them. [R' Eliezer] said to him: Go back and tell them: Do not fret about your voting. Thus I have received from R' Yochanan ben Zakkai, who heard it from his teacher, and his teacher from his teacher: A legal tradition to Moshe from Sinai that Ammon and Moav must give the tithe of the poor in the seventh year.

R' Eliezer speaks out against the sages of Yavneh, who 'innovate innovations', while he possesses the Halacha from the times of Moses at Sinai. This is a result of ignorance. If the sages were in need of the tradition that he knows, they did not need to have to debate at all.

Historically, and from parallel fragments, it is clear that the fragment from Tractate Chagiga deals with the revolution in Yavneh on 'that day'. The opposing beliefs of R' Eliezer and R' Yehoshua are here shown reflected in their assessment of what happened in Yavneh of R' Elazar ben Azaryah.

Let us now return to Tractate Avot, chapter I, 1-2 (cited above). Receipt and Transmission of the Torah is described only until the times of R' Yochanan ben Zakkai. R' Eliezer is characterised as a 'cemented cistern that loses not a drop', but eventually he is not the receiver from R' Yochanan ben Zakkai, but R' Yehoshua. He is not presented as a 'receiver' or part of a couple. The study hall is now wide open for everybody, because people are judged by the contents of what he says and not what he is.

In the generation of R' Yochanan ben Zakkai the split between the disciples of Hillel and Shammai took place, threatening the Torah and the people. The disciples of R' Yochanan ben Zakkai saved the situation by defining new ways of debating and decision making. This is the Yavneh revolution that happened on 'that day'.

The continuation is not linear anymore. The Torah transmitted from Yavneh in the following generations is a combination of the 'Law of Tradition' and the 'Law of Debate'. The process ended in a dialectic synthesis: The two extreme viewpoints were united in one comprehensive whole.

This is illustrated in the fragments in Tractate Sanhedrin 68 and 101. They contain parallel descriptions (with some important distinctions) of the visit of the disciples of R' Eliezer the Great on the day of his death.

In the Mishna (Tractate Sanhedrin 67a) R' Akiva learns the law of 'two [people] gathering cucumbers' from R' Eliezer. In the Gemara (Tractate Sanhedrin 68a) he receives the law as a tradition, but afterwards he asks R' Yehoshua for an explanation of the law, and only then does he accept it: R' Akiva considers both R' Eliezer and R' Yehoshua as his teachers. Also, in Tractate Sanhedrin, 101a, R' Eliezer remarks that R' Akiva is the only one who asks for his opinion, i.e. inquires about the tradition.

R' Akiva is the leader of the synthetic method. His style is a combination of tradition (which he learned from R' Eliezer) and debate (which he learned from R' Yehoshua). This approach continues in the following generations. Therefore R' Akiva is considered the father of the entire Oral Law as it has come down to us. The editor of the Jerusalem Talmud, R' Yochanan famously expressed this in the following way (Tractate Sanhedrin 86a):

> Stam Mishna [an anonymous passage in the Mishna is attributed

> to] Rabbi Meier, Stam Tosefta[24] R' Nechemiah, Stam Sifra[25] R' Yehudah, Stam Sifri[26] R' Shimon, and all of them according [to what they had learned from] Rabbi Akiva

1.5 The Understanding of Rules

So far we have examined the question of authority and adherence to binding precedents, but the Talmudic anarchy also expresses itself in other ways. One of the most salient expressions of this is the following example from Mishna in Tractate Eruvin, 26b (it is continued in our second book[41]):

> We may make an eruvei[27] [techumim] and a shitufei[28] [mevoot] with all [types of food] except for water and salt. And all [types of food] may be purchased with maaser[29] [sheni] funds except for water and salt.

This means that it is allowed to use money from maaser sheni or make an eruv with all foodstuffs, except water and salt. It seems this is a very precise definition, and one would not expect that other foodstuffs would also not be allowed to buy from maaser sheni. But the Gemara immediately brings the following saying (Tractate Eruvin, 27a):

R'Yochanan said: We cannot learn [i.e. extract categorical rulings] from general rules, and even where [the rule concludes] by saying 'except'

The Gemara brings an anarchistic rule (!), which states that we cannot learn anything from rules, even if they are specific, i.e., they itemise the exceptions. Indeed, the continuation of the fragment lists other foodstuffs that one cannot buy with maaser sheni money.

The Gemara later adds further rules to the anarchistic rule (Tractate Eruvin, 27a):

> Since [R'Yochanan] said, Even where it says 'except', it is implied that the statement does not refer to here [our Mishna]. To where does it refer? It refers to there: "All positive mitzvoth that are time-bound, men are obligated [to perform them] and women are exempted, and those which are not time-bound, both

[24]The Tosefta (Aramaic: Additions, Supplements) is a compilation of the Jewish oral law from the period of the Mishnah .In many ways, it acts as a supplement to the Mishnah

[25]Sifra is the Halakic Midrash (classical Jewish legal Biblical exegesis), based on the biblical book of Leviticus.

[26]Sifri refers to either of two works of Halachic Midrash (classical Jewish legal Biblical exegesis), based on the biblical books of Bamidbar (Numbers) and Devarim (Deuteronomy).

[27]An Eruv is a ritual enclosure around most Orthodox Jewish and Conservative Jewish homes or communities

[28]A Shituf Mevo'ot is similar to an eruv

[29]The Maaser Sheni, meaning *Second Tithe* in Hebrew, is a tithing practice in Orthodox Judaism with roots in the Hebrew Bible

> men and women are obligated". And is this an [absolute] rule, that all positive mitzvoth that are time-bound women are exempted from performing? But there are [the positive mitzvoth of] matzah, happiness [during festivals], and assemblage, which are all positive mitzvoth that are time-bound – and [yet] women are obliged.
>
> And [is it true that] all positive mitzvoth which are not time-bound, women are obliged? But there are Torah study, being fruitful and multiplying, and redeeming a [firstborn] son, which are [all] positive mitzvoth that are not time-bound, and women are [nevertheless] absolved [from these obligations]. So R'Yochanan said: We cannot learn [categorical rulings] from general rules, and even where [the rule] states 'except'.

Thus, the exemption of women from time-bound obligations is a rule that should be examined carefully. It should not be taken too seriously. At the end of the fragment yet another non-obligatory rule is stated (Tractate Eruvin, 27a):

> Abayeh said, and some say R' Yirmiyah: We also learned this in a Mishna: also another rule was said about the laws of a zav:
>
> "All the things that are borne upon a zav are tamei, and all things which a zav is borne upon are tahor, except for [things which are] suitable for reclining or sitting [upon], and a person".
>
> And there are no more [exceptions]? But there is an object used for riding. This object for riding what is it like? If he sits on it, it is like sitting. We meant to say thus: There is the upper part of a saddle, [why was it not mentioned?] For it was taught in a Braita[30]: The saddle is subject to the tumah of moshav, and the pommel is subject to the tumah of merkav. So we learn from this, [that] we cannot learn [by making deductions] from general rules, and even in a place where it says 'except'.

The fragment ends by returning to the Mishna from above:

> Ravina said, and some say it was Rav Nachman: We also learned thus in our Mishna: With all [types of food] we may make an eruv [techumin] and a shituf [mevoot], except with water and salt. And there are no more exceptions? But there are truffles and mushrooms [also disqualified, but not mentioned]? So we learn from this, that we cannot learn [halachot by making deductions] from general rules, and even in a place where it says 'except'.

[30]Braita (pl: Braitot) refers to a statement or passage found in the Talmud that could have been included in the Mishna, but is nowhere to be found there.

As observed above, other foodstuffs not to be bought from maaser sheni money are here mentioned, despite the fact that the Mishna, supposedly, specified all excluded items.

The structure of the Talmud is casuistic. It seldom defines rules, and even when it does, it assigns them only limited warranty. The Talmud allows itself to change the interpretation of the Mishna and other Tannaic sources from the literal meaning, according to the judgement of the Amoraim. It distorts the words of the Mishna and the Braitot in such a way, that the declared obligation to these sources becomes almost ridiculous. There are cases where it is explained that a certain Mishna deals with a special case ('okimta'), or missing sentences are added ('chasurei mechasrei'). All this is done in order to make the Mishna fit to logic and what is reflected in parallel sources. The conclusion is that the relationship of the Halacha to obligatory texts and stringent rules is weak. The Halacha and the Talmud do not like rules, and when such rules do appear, they are of bounded status.

It is possible that the reason for formulating rules at all was the need to conserve knowledge passed orally. It is forbidden to write the Oral Torah down (see Tractate Gittin, 60a). This probably shows the wish to leave it open to interpretations and applications. At different times in history the sages decided to diverge from this prohibition, and write the information down in order not to forget it. This happened when the amount of knowledge became too big, and when the convulsions of the Diaspora threatened the capability of the collective memory to store all the oral knowledge. As a part of this aim the rules were created. The purpose was more to safeguard the knowledge, rather than a directive to Poskim (the practical deciders of the Halacha). Perhaps this is the reason for the contempt in which the rules are held.

1.6 Autonomy and Authority in Halachic Decision Making

The anarchism of the Halacha is also seen in the autonomy that the Halacha gives the Posek. We shall examine some post-Talmudic expressions of this, found in the lack of obligation to rules, and the lack of obligation to precedents and decisions by previous generations.

This subject was extensively discussed by Rabbi Asher[31]. He deals with the question whether the Gaonim[11], who came after the Talmud, have authority over the following generations (the Rishonim[8]). In Piskei haRosh[32] on Tractate Sanhedrin 84, section 6) he brings two major opin-

[31]Asher ben Jehiel (or Asher ben Yechiel, sometimes Asheri) (1250 or 1259 – 1327) was an eminent rabbi and Talmudist best known for his abstract of Talmudic law. He is often referred to as Rabbenu Asher, "our Rabbi Asher" or by the Hebrew acronym for this title, the ROSH (literally "Head").

[32]Piskei haRosh is a summary of the Halacha derived from the Rosh's Talmudic commentary, compiled by his son.

ions: The Raavad[33] says that whoever disagrees with the Gaonim, errors in the Mishna, and to diverge from an obligatory canonical source has serious personal implications for the Dayan (judge). The Baal haMaor[34] says that whoever disagrees with the Gaonim, errors in his reasoning. They both agree that it is an error to disagree with the Gaonim. The question is the status of this error.

But the Rosh himself states that the judge has all the right to make Halachic decisions according to what he thinks, without building on precedents. Exceptional to this is the Talmud, which, as we have explained above, is an obligatory canon.

The Remah[13] establishes the law in Choshen Mishpat[35], section 25. He asserts that the Acharonim may disagree with the Rishonim, and even we are allowed not to accept their rulings, if there is a good reason for it.

One should keep in mind that this was written at a time, when autonomous decision making was on the wane. It appears in the Shulchan Aruch, despite the fact that the two authors are strong advocates of precedent based rulings. Much controversy arose at the beginning of modern times (the Acharonim) because of the change in direction (the codification controversy).

The Maharal[36] writes in Netivoth Olam ("Pathways of the World" — a work of ethics) that autonomous decision making has a most important value, even at the price of erroneous judgments. He states that God prefers the one who decides by his intelligence and not by precedent. Even if a judge makes a mistake, he is to be preferred to the one who judges according to a 'book' (precedent), who may be right.

On the other hand, Ri Megas[37] was asked whether it is allowed to let somebody decide Halacha according to the Gaonim, even if he does not know the Talmudic source and procedure. His answer (responsum 114) contradicts the opinion of Maharal: It is better that somebody decides

---

[33]Rabbi Abraham Ben David, The Raavad (1125-1198), born in Posquieres, Provence, France. He was a a great commentator on the Talmud, Sefer Halachot of Rabbi Yitzhak Alfasi and Mishne Torah of Maimonides.

[34]Zerachiah ben Isaac Ha-Levi Gerondi (called the Baal Ha-Maor — author of the book Ha-Maor) was born about 1125 in the town of Girona, Spain and died after 1186 in Lunel. He was a famous rabbi, Torah and Talmud commentator and a poet.

[35]Choshen Mishpat (Hebrew for "Breastplate of Judgement"). The term is associated with one of the four sections of Shulchan Aruch. This section treats aspects of Jewish law pertinent to finance, torts, legal procedure and loans and interest in Judaism.

[36]Judah Loew ben Bezalel, (c. 1520 – 17 September 1609) known as the Maharal of Prague, or simply The MaHaRaL, the Hebrew acronym of "Moreinu Ha-Rav Loew," ("Our Teacher, Rabbi Loew"). He was an important Talmudic scholar, Jewish mystic, and philosopher who served as a leading rabbi in the city of Prague in Bohemia for most of his life.

[37]Joseph ben Meir ibn Megas or Megas (1077–1141) was a Rabbi, Posek, and Rosh Yeshiva in Lucena. He is also known as Ri Megas, the Hebrew acronym for "Rabbi Joseph Megas".

according to the 'book', even if he does not fully understand (for he will usually reach the truth), than if he were to decide according to reason (which may be wrong).

It would seem that the disagreement is about whether there exists a single Halachic truth, or whether what the Dayan and Posek gives as a reasoned decision becomes the Halachic truth. However, if we examine the two sources carefully, we see that they actually agree in principle. The Maharal does not assert that there is no Halachic truth. He speaks about somebody who makes an error in his judgement (but he considers this preferable to the Posek who goes according to the 'book'). The Maharal believes that autonomous decision making has an intrinsic value, and this is sometimes better, even if in error. This is similar to what the Rosh had to say.

On the other hand, the Maharal does not conclude that just anybody can make such decisions. He ends the enquiry by saying that in his generation not many are capable of this. He also limits the recommendation of Halachic autonomy only to those who are really competent ('Bar Hachi' in the words of the Rosh)..

Examining the approach by Ri Megas we observe the same elements. The Ri speaks about the danger of a Halachic error, and is not prepared to let just any Posek make independent decisions. On the other hand it is clear from his words, that if somebody is indeed competent ('Bar Hachi'), he may decide according to his intelligence. But in the estimate of the Ri only few of his generation are qualified.

So the Maharal and Ri Megas actually say the same thing, and they only differ in their estimate of the factual situation: Are there or are there not competent ('Bar Hachi') people in their generation.

We see that there is indeed a place for Halachic competence and precedents, but only in a very limited manner. But apart from the discussion itself about this matter among the sages, even the most conservative among them do not believe in an absolute attachment to old sources.

The feeling of continuity that accompanies the Halachic study and debate is complemented by a sense of autonomy. The Posek feels obliged to express his private opinion and fight for it, even if it does not fully overlap the Godly truth. This is illustrated in the words of Rabbi Kook[38] on a contradiction between two Talmudic sources that both characterise R' Eliezer the Great. In one source (Tractate Sukkah, 28a) R' Eliezer states that he has never said anything he had not heard previously from his rabbi (see also Tractate Yomah, 66b and other places). The other source is in Avot d'Rabbi Natan[39] (86), where it is stated that R' Eliezer said things that 'had never been heard

[38] Abraham Isaac Kook (1865–1935) was the first Ashkenazi chief rabbi of the British Mandate for Palestine, the founder of the Religious Zionist Yeshiva Merkaz HaRav, Jewish thinker, Halachist, Kabbalist and a renowned Torah scholar.

[39] Avot de-Rabbi Nathan, usually printed together with the minor tractates of the Talmud, is a Jewish aggadic work probably compiled in the geonic era (c.700–900 CE).

before'. Rabbi Kook explains that R' Eliezer in the second source did not say that 'these were things he had never heard' — which would indeed be a contradiction — but 'things that had never been heard before'. This means that his rabbi, R' Yochanan ben Zakkai actually said those things, but only R' Eliezer heard them, while other listeners did not.

Rabbi Kook here describes the feeling of innovation arising from continuity. On one hand the sage only discloses the tradition he has received, but this disclosure expands and generalises the tradition. It is full of novelties and new directions. This expresses the dichotomy in Halachic discourse. On one hand innovation is recognised and even encouraged, on the other hand continuity and convention are stressed. As we have seen, the attachment to custom is very flexible, and sometimes innovation is greater than tradition.

1.7 Should One Expect that there Exists a Logic of the Talmud?

So far we have seen a picture of a normative system, which varies over time, and goes through several improvements. It does not adopt strict rules or a rigid framework. So the question arises whether there is any purpose in examining the logic of such a system? Or rather, whether a specific logic actually exists.

Let us first remark, that in a system where the debate is more important than the conclusions, one would expect a rich logic worthy of examination and analysis. Were the Halachic canon just a collection of laws, or even directions and values, perhaps there would not be any expectations of a clear logical basis. As we have shown in our fourth book[40] of the series, legal systems in general do not seek to apply a specific logic, and the logical research relating to such systems is minimal. One reason for that is that their development is not based on logical deduction and debating rules. They aim at specific purposes and the methodical formulation of agreements and social frameworks. The Halacha, on the other hand, is based on debate and the logical conclusions that arise from this. Hence, contrary to expectation, it would seem that there is a place for logical research.

A difficult problem of methodology arises here. The number of sources, periods and sages is large, and the question is whether there exists a Halachic or Talmudic logic as a unique category enabling separate research. In modern Talmudic research it is assumed that the development is unmethodical, the result of different cultural and intellectual directions and of varying social and environmental pressures. This is why the Talmudic research, in contrast to classical study of the Talmud, does not aim at harmonising among various Halachic sources. It considers each source as independent, and will at the most compare them.

[40] Abraham M., Belfer I., Gabbay D., Schild U., *Temporal Logic in the Talmud* (in Hebrew), College Publications, King's College, London, UK, 2011

On the other hand, as we have explained in our second book[41], we do not accept the situation described above. The relevant logic is indeed developing through the ages, but our historical and methodological assumption is that this logic is disciplined and consistent. Over time it becomes clearer that this is indeed one logic. Part of it is universal — and this is the part most interesting for us — and part of it is unique to the Talmud and to Halacha.

Our assumption is that examination of the later stages of development cast light upon the earlier stages. We believe that the way to understand the significance of the Talmudic-Halachic debate is through the prism of its later stages. This is also the assumption of the traditional scholar, but one should not be surprised to discover that only seldom has methodical research been based on this assumption.

In order to base our assertion we observe that despite the anarchistic picture described above, there is a continuous historic process, which seems entirely opposed to anarchy. It is a transition to causal and associative thinking, the use of rules and strengthening of methodology in the Talmudic and Halachic thinking. Talmudic research sees this as a later development. However, we suggest that one here sees a germination of seeds previously sown. Within the historical process each generation of sages decode the principles that form the basis of tradition received from previous generations. They begin to use such rules as more rigid rules of interpretation. The concepts crystallise, become formalised and canonised. They are now rules of logic in some sense.

The significance of this has been described in detail in our second book[41] in the series, and will not be repeated here. The main assertion is that the Talmudic ways of thinking does not change, but become more general, methodical and logical. In later stages earlier types of thinking become systematic. In the second book we have shown that the rule of Klal uPrat, which was one single rule in the times of Hillel the Elder, became three or four rules in the list of R' Yishmael. This what has happened in general to Halachic thinking.

In this sense, our investigations form a continuation of this Halachic tradition. We too are attempting to discover the ways of reasoning of our predecessors, to conceptualise and give them a foundation, and analyse them using modern logical tools. The results so far are very encouraging, and show that this process has great opportunities to persist.

1.8 Two Important Distinctions

Before concluding we have to make two important distinctions between Talmudic Logic and Mathematical Logic. First of all, Talmudic Logic has

[41] Abraham M., Gabbay D., Hazut G., Maruvka Y.E., Schild U., *The Textual Inference Rules Klal uPrat* (in Hebrew), College Publications, King's College, London, UK, 2010

more intuitive characteristics. Prima facie it does not seem to be formalistic, while Mathematical Logic and Natural Science in general are based on formal thinking.

This distinction may be understood at two different levels: Essence and Action. Talmudic thinking does not apply formal rules, at least not rules that have been explicitly formulated. That does not mean that such rules do not exist, as we have seen above — later stages of Talmudic thinking conceptualise rules that are based on earlier Tamudic thinking. But there is also an essential difference between standard logic and Talmudic thinking. Standard logic deals mainly with necessary and certain inferences, i.e., deduction. All that is not part of such inferences are not part of classical logic. Logic also considers other types of inference (induction, abduction and analogy), but does not provide a methodical and formal foundation for these types. The Talmud, on the other hand, is almost totally based on uncertain inferences.

Nachmanides[42], in his Milhamot Hashem (Wars of the Lord) defends the decisions of Alfasi[43] against the criticisms of Zerachiah ha-Levi of Girona. He explains that in the Talmud and in the Talmudic debate there are no absolute statements like in Mathematics, and not even empirical evidence like in Physics. There are disagreements about interpretation and what counts. There are not absolute logical proofs, but a criterion of what is or is not reasonable. This is Nachmanides' characterisation of the Talmudic debate.

All fields of knowledge and science, except Logic and Mathematics, belong to the category of domains with uncertain conclusions, like the Talmud and the Halacha. Hence, traditionally, inferences in those areas are considered outside standard logic. However, as we have shown in our books and papers, this is not true. Also uncertain inferences (like analogy and induction) may be formalised, and this is relevant to all science and human knowledge.

Rabbi David Cohen (the Nazirite Rabbi[44]), the great disciple of Rabbi Kook[32], devoted his book: *Kol Nevu'ah*, to this idea. He places two kinds of thought against each other. One is the Greek (scholastic) logic, which is material-visual, the other is the Jewish logic (Talmudic-Halachic), which is spiritual-acoustic. The Jewish philosophy is not single valued as the classical logic, but is based on sound and deep understanding. There is a preference

[42]Nahmanides, also known as Rabbi Moses ben Nahman Girondi, Bonastruc ça Porta and by his acronym Ramban, (Girona, 1194 -=- Land of Israel, 1270), was a leading medieval Jewish scholar, Catalan rabbi, philosopher, physician, kabbalist, and biblical commentator.

[43]Isaac ben Jacob Alfasi ha-Cohen (1013–1103) — also known as the Alfasi or by his Hebrew acronym Rif (Rabbi Isaac al-Fasi), was a Talmudist and Posek (decider in matters of Halacha)

[44]David Cohen (1887 – 1972) (also known as "Rav Ha-Nazir", The Nazirite Rabbi) was a rabbi, talmudist, philosopher, and kabbalist. A noted Jewish ascetic, he took a Nazirite vow after making aliyah to Israel.

of the reasonable over the less reasonable, and of what is heard over what is not heard well (even without sharp evidence, as also found in the words of Nachmanides).

The Nazirite Rabbi also asserts that the Halachic inferences are the foundation for this alternative logic, like the basic deductions in Aristotelian Logic. The Talmud prefers analogy and induction, i.e., softer inferences, in order to reach conclusions. The Talmud does not draw back from conclusions like 'perhaps it is also possible otherwise?', nor from the lack of certainty of the conclusions. It weighs the alternatives against each other, but is prepared to reach a decision where decision is not incontestable. Also obligations relating to doubts have a central place in Halachic thought, and there is an entire system of rules of decision and behaviour in the case of doubtful situations.

The Talmud and the Halacha live in an atmosphere where doubt is present at all times. They defer sharp conclusions, and prefer decisions based on preferences. The Talmud deals with life in all its complexities, and does not deal with abstract ideas. Its approach is usually casuistic, i.e., it will consider a concrete case and not the abstract idea itself. Generalisations arise from the consideration of the specific case, and do not precede it. Here too one sees the inductivity of Talmudic thought, which goes from the special case to the general one, and prefers this method to deductive thought, which goes from the general to the special case.

2 OBLIGATIONS AND PROHIBITIONS IN TALMUDIC DEONTIC LOGIC

2.1 Introduction

This chapter examines the deontic logic of the Talmud. We shall find, by looking at examples, that at first approximation we need deontic logic with several connectives:

$O_T A$	Talmudic obligation
$F_T A$	Talmudic prohibition
$F_D A$	Standard deontic prohibition
$O_D A$	Standard deontic obligation.

In classical logic one would have expected that deontic obligation O_D is definable by

- $O_D A \equiv F_D \neg A$

and that O_T and F_T are connected by

- $O_T A \equiv F_T \neg A$

This is not the case in the Talmud for the T (Talmudic) operators, though it does hold for the D operators. We must change our underlying logic. We have to regard $\{O_T, F_T\}$ and $\{O_D, F_D\}$ as two sets of operators , where O_T and F_T are independent of one another and where we have some connections between the two sets.

We shall list the types of obligation patterns appearing in the Talmud and develop an intuitionistic deontic logic to accommodate them. We shall compare Talmudic deontic logic with modern deontic logic.

2.2 Motivating Talmudic deontic logic TDL

This chapter is written for researchers in Deontic Logic and Contrary to Duties who would like to know how things stand in Talmudic logic. It is an expanded version of [5]. To set the scene for this chapter, we give some short background material.

The simplest and historically first logical system offered for dealing with obligation is Standard Deontic Logic **SDL**, which is the modal logic **KD** for an operator $O_D A$ reading 'A is obligatory'. The semantics for O are models of the form (S, R, h), where $R \subseteq S^2, h$ is the assignment to the atoms, assigning each atom q of the language a subset $h(q) \subseteq S$, and R satisfies $\forall x \exists y x R y$.

This system was too simple and researchers in the community offered systems with dyadic modalities $O_D(A/C)$, reading 'A is obligatory in the context C'. This was a response to contrary to duty examples which could not be properly modelled by the unary O_D.

One such famous example is the Chisholm set:[45]

1. It ought to be that a certain man goes to assist his neighbour.
2. It ought to be that if he does go he tells him he is coming.
3. If he does not go he ought not to tell him he is coming.
4. He does not go.

If we use H for 'help' and T for 'tell', we have two options to formalise this set, either with $O_D X$ (unary) or with $O_D(X/Y)$ dyadic.

clause	monadic	dyadic
1.	$O_D H$	$O_D(H/\top)$
2.	$H \to O_D T$	$H \to O_D(T/H)$
3.	$\neg H \to O_D \neg T$	$\neg H \to O_D(\neg T/\neg H)$
4.	$\neg H$	$\neg H$

[45]The translation of (1)–(4) must give four consistent and logically independent sentences adequately representing the linguistic text.

The following sums up the spirit of the research of the deontic community.

1. Find reasonable logical systems involving various monadic or dyadic modal operators with possible world or preferential semantics in which various linguistic deontic sets can be consistently and adequately formalised.

2. Emphasise the CTD examples and calibrate your logics to deal with various problems associated with them.

The community lays stress on the theory of CTDs as distinctly characteristic to deontic logic, which sets it apart from being a secondary applied branch of modal logic. It is also felt that the essence of the deontic area is the possibility of violations and hence the core of deontic logic as a discipline distinct from modal logic is its theory of CTD.

For our purpose a contrary to duty system is a set Δ of formulas of the form $\{\delta_1, \ldots, \delta_n\}$ where

$$\delta_i = O(X_i / Y_1 \wedge \ldots \wedge Y_{k(i)}).$$

Given a consistent set

$$\theta = (E_1, \ldots, E_k)$$

we consider the set

$$\Delta_\theta = \{X_i | \delta_i(X_i / Y_1 \wedge \ldots Y_{k(i)}) \in \Delta \text{ and } \theta \vdash Y_j, j = 1, \ldots, k(i)\}$$

Δ_θ is the set of obligations triggered by the context θ. Δ_θ may be an inconsistent set and part of any CTD logic is to "recommend" a consistent subset $\Delta_\theta^{\text{con}} \subseteq \Delta_\theta$. The "logic" has to deal coherently and in a compatible manner with common sense with the relationship between pairs of the form $(\theta, \Delta_\theta^{\text{con}})$ and $(\theta', \Delta_{\theta'}^{\text{con}})$. As far as we know, no comprehensive theory of this form exists. See references [93; 94; 58; 59; 28; 63].

In contrast with the above, The Talmud, being a religious code of law, given to us by God in the Bible, has two types of deontic rules: action obligations and action prohibitions. Both types represent the will of God for us to obey. This is why at a first logical approximation we need two independent deontic operations O_T and F_T (the subscript 'T' stands for 'Talmudic') as well as the standard deontic Obligation O_D and prohibition F_D.

There are some points we need to make clear. The variables X that go into the connectives $O_T X, F_T X, O_D X$ and $F_D X$ denote actions like work, lift, steal, wear Tefilin (Tefilin is something men wear when they offer morning prayers during week days), etc. and not lack of action like resting, not stealing, etc. When we negate them and write $\neg X$, we denote lack of action.

We are not going to discuss how to determine what is considered action and what is to be considered inaction. This is a separate issue. We assume it is always clear, for any candidate formulas A and $\neg A$, which is the action formula and which is the inaction formula.

One might think that we can model obligations and prohibitions using only one deontic operator O, letting OX represent obligations and $O\neg X$ represent prohibitions. However this is not correct. Our obligations and prohibitions can apply either to X or to $\neg X$. See examples below under the heading "Type 3: Strong obligation/prohibition'. So $O_T X$ is a Biblical obligation to take action X. $O_T\neg X$ is a Biblical obligation not to take action X. $F_T X$ is a Biblical prohibition to take action X and $F_T\neg X$ is a Biblical prohibition not to take action X (i.e. we are prohibited from choosing not to take action X). So $O_T X$ is not equivalent to $F_T\neg X$. So if $X =$ wear Tefilin, then having an obligation to wear it is not the same as being prohibited from not wearing it. So in some cases God requires us to obey both i.e. $O_T X \wedge F_T\neg X$. The reader should recall intuitionistic logic where $\neg\neg A$ is weaker than A, so the negation used in these commands have intuitionistic flavour. (In fact, the Talmudic system will be modelled in intuitionistic modal logic.)[46]

If we look at this situation as logicians, we can say we have here three pairs of modal operators, each pair being of the form (Necessity of the form NX and Possibility of the form $PX = \neg N\neg X$). The pairs are $(N_i, P_i), i = 1, 2, 3$ as follows.

1. $O_T X$ and $\neg O_T\neg X$
2. $F_T\neg X$ and $\neg F_T X$
3. $O_D X$ and $\neg F_D X = \neg O_D\neg X$.

Since the Talmud gives no connections between O_T and F_T, we have to represent them as two pairs $\{N_1 X = O_T X, P_1 X = \neg O_T\neg X\}$, and $\{N_2 X = F_T\neg X, P_2 X = \neg F_T X\}$.

[46]In Talmudic logic we have that $\neg\neg F_T(A)$ is not equivalent to $F_T(A)$. The first is only a weak prohibition, a recommendation for good behaviour in the eyes of God, while the second is a full fledged strong prohibition. This is reflected in our use of intuitionistic logic as a basis.

The perceptive reader might say that perhaps we could obtain a similar result without the use of intuitionistic logic, by considering explicit permissions which are distinct from the negation of a prohibition.

More specifically, we introduce an additional modal operator P, with the axiom $F \to \neg P$, but without the axiom $\neg P \to F$. In that case, the negation of permission may correspond to a weak prohibition, but without requiring intuitionistic logic for this purpose.

However introducing another independent operator is too strong and does not manifest the intention that $\neg\neg F_T(A)$ is only a recommendation of $F_T(A)$. Furthermore the idea of explicit permissions is not compatible with Talmudic thinking. God never said in the Bible "you are allowed to do this". He only delivered to us Obligations and Prohibitions. See Section 5.1 for further discussion.

This can be made clearer when we consider the operational differences between $O_T A$ and $F_T A$ and $F_D A$.

1. If you obey $O_T A$ then God rewards you. You are also obliged to spend 20% of your income to enable yourself to fulfil your obligation.

2. If you violate $F_T A$, and actually do the forbidden A, then you will be punished (by God and or by law/society). Also you should devote 100% of your income to enable yourself to avoid doing A.

Therefore for the same X, if the Bible says $O_T X$ then 1. applies and if the Bible says $F_T \neg X$, which in practice means the same to us, then 2. applies.[47]

$F_D A$ says it is forbidden to have A for whatever reason, without going into the fine tuning of why this is so. It may arise from a Biblical $O_T \neg A$, or from $F_T A$ or from some related $F_T Y$ or whatever.

For example, in Type 1A: Obligation with deontic prohibition below we have O_T (wear Tefilin during prayer). From this it follows that F_D (pray without wearing Tefilin).

However we do not have a direct Biblical prohibition F_T (pray without Tefilin), and therefore if one actually does pray without wearing Tefilin, there is no punishment from God.

Note that we do not necessarily have any connections like

$$O_T X \rightarrow \neg F_D X$$

and

$$F_T X \rightarrow F_D X.$$

If we had them we could have derived

$$O_T X \rightarrow \neg F_T X.$$

However we know that there is no such axiomatic connection in Talmudic logic. The reason for that is as we mentioned earlier, O_T and F_T are in general generic and possibly conflicting, and it is the Rabbis who decide day-to-day how to apply the commands in any given situation.

It is possible also to have both $F_T X$ and $O_T X$ for the same X (even though on the surface this seems contradictory) because X may be a generic kind of predicate and it is expected that the Rabbis will decide for each situation $\mathbf{s}$ which obligation/prohibition applies. In fact, in many cases the Bible gives recipes (more precisely there are indirect hints in the Biblical text but the main derivation of recipes is done in the Talmud) for making

[47] A main difference between biblical obligation and prohibition and ordinary traditional Deontic obligation and prohibition is that violation of a biblical prohibition would imply a sanction whereas fulfilment of the corresponding biblical obligation implies a reward. This in an important dimension, and will be further discussed in Section 5.1

such decisions. In our model these recipes are part of the (nonmonotonic) mechanisms of conflict resolution.

It is the job of the Rabbis to make decisions (according to some principles) how to resolve conflicts between obligations and prohibitions when applied to any particular situations.

The emphasis of Talmudic Deontic Logic is therefore on

1. Deciding what are the Biblical $O_T X, F_T X$. (This has been done: there are 613 master ones, though opinions differ as to which are included among these 613.)
2. Deciding which Biblical $O_T X, F_T Y$ apply to any new arising situation $\mathbf{s}$.
3. Resolving possible conflicts between applicable rules for any $\mathbf{s}$.

The role of CTDs is not central to the Talmudic system, nor is the theoretical maintainance of consistency. The Biblical rules are known to cause conflict and established procedures and recommendations and institutions for conflict resolution and practical day-to-day decision making are also given by the Bible.

Note that there are differences between this decision making process and precedents and legislation in law. We shall not go into that here. See, however, Section 5.3.

The following table, Table 1, compares Talmudic Deontic ideas with their modern counterparts.

To compare CTDs, let us look at some examples from the Bible.[48]

EXAMPLE 1 (Chisholm variant 1).

1. You ought to have a ceremonial meal during the Passover festival.
2. If you have your meal you ought to say prayer (blessing, grace).
3. If you do not have the meal you ought not say the prayer (blessing).
4. You do not have the meal.

(1)–(3) are Biblical obligations. We formalise them using dyadic modalities.

1. $O_T M$ (or $O_T(M/\top)$)
2. $M \to O_T B$ (or $M \to O_T(B/M)$)
3. $\neg M \to F_T(B/\neg M)$
4. $\neg M$.

[48]The Talmud interprets the Bible. So when we say Talmudic logic, this includes Biblical logic.

	Deontic community	Talmud	Comments
Sources of obligations and prohibitions	common sense, law, moral code	Bible/God	It took hundreds of years to study and summarise the Talmudic obligations and prohibitions. 613 major types were finally agreed upon by the end of the middle ages, though as we already mentioned, opinions differ as to which are included among these 613.
Formalisation	Monadic or dyadic operators, preference or possible world models.	Two levels O_T, F_T, and O_D, F_D. The handling metalogic is some kind of time action logic	Modern deontic logic is a well developed area. This chapter is a first attempt in formalising Talmudic deontic logic
Status of CTD	central	marginal	The Talmud views CTD as just more conditional obligations
Conflict resolution	Recognised but not central yet. The community is beginning to address the problem.	central	Deontic community recognises the problem of inherited conflicting CTDs. They emphasise consistency. Talmud expects inconsistency even of original obligations. Emphasises methods of resolving conflicts.
Status of violations	Violations are expected, that is why CTDs are central, but there is no reward for obeying a CTD.	Obeyance is expected	Talmud emphasises punishment for violations and reward for obeyance.

Table 1.

Note that the Bible is explicit about $F_T(B/\neg M)$ and does not say $O_T(\neg B/\neg M)$. The Bible says generally "Do not use the name of God in vain", which applies to this case as well!

> Exodus 20:7
> You shall not take the name of the LORD your God in vain, for the LORD will not leave him unpunished who takes His name in vain.

We do *not* have the equivalence $O_T(\neg x/z) \equiv F_T(x/z)$.

Compare the above with the following.

EXAMPLE 2 (Chisholm variant 2).

1. We are obliged to eat meat from sheep at passover.

   > Exodus 12:21
   > Then Moses called for all the elders of Israel and said to them, Go and take for yourselves lambs according to your families, and slay the Passover lamb.

2. If we eat meat we should slaughter the sheep humanely.
3. If we do not eat meat we should not slaughter the sheep.
4. We do not eat meat.

The translation is as follows (E is Eat and H is sheep):

1. $O_T(E/\top)$.
2. $E \to O_T(H/E)$.
3. $\neg E \to F_D(H/\neg E)$
4. $\neg E$

Note that in (3) we used F_D because the Bible is not explicitly prohibiting killing animals for no reason but the prohibition follows from Rabbinical practical rulings.

Thus the reward from God for obedience is different in the two cases. Note that it is easier to avoid the Chisholm paradox for examples 1.1 and 1.2 since our logic language is more refined.

The rest of this section will give examples of the major existing types of Talmudic obligations and prohibitions and formalise the examples in terms of O_T, F_T and F_D. The reader should note that we may have less or different paradoxes for the Talmudic system, which has more operators and so more fine distinctions can be made. Furthermore if in ordinary deontic logic we allow more operators to stand for strong moral (parallel to Talmudic) obligations and prohibitions, then we might find that some paradoxes disappear. Although we have not given yet to the reader the axiom system and semantics for these operators, we have given enough of their intuitive meaning and this should suffice for our initial formalisation.

Let us now briefly describe the eight types of obligations and prohibitions available in the Talmud.

We shall also give a preliminary intuitive formalisation in terms of O_T, F_T and F_D (note that O_D is definable from F_D, so we do not need it). In the sequel, we distinguish Types 1A, 1B and 1C. They all arise from the same Biblical Talmudic obligation O_T. The differences between them is practical implementations, as summarised in Table 2.

Type 1A. Obligation with deontic prohibition

As an example, we have to respect and honour our parents (this is one of the Ten Commandments), so we have O_T (Respect Parents). If we do not respect our parents, there is a violation. See Table 2 item 1A. The Bible says respect your parents so that you will live long and prosper. It does not threaten punishment if you do not.

> Deuteronomy 20:8
> Honor thy father and thy mother, that thy days may be long upon the land which the LORD thy God giveth thee.

Perhaps a modern example will help. We all read some Harry Potter books. The newspapers reported that the author J. K. Rowling gave her father copies of the first edition of her books, signed and dedicated by her. The idea was that he was supposed to keep them. The father needed money and so he sold them. We formalise the intention/convention by $O_T\text{keep} \wedge F_D\neg\text{keep}$.

He is not supposed to sell them because he is expected to keep them.

Type 1B. Weak obligation

There is an obligation to live in the land of Israel. The question is whether from this obligation there is a deontic prohibition on living outside Israel. The answer is no, according to a minority opinion. Now if you do not live in Israel, there is no violation. See Table 2, item 1B. This is a unique case where the weak obligation is some sort of recommendation. You get a reward if you do it but there is no violation if you do not do it.

Type 1C. Prohibition arising from positive obligation

We need to let the land rest every seven years. As part of this the fruits of trees on the seventh year are allowed to be eaten by anyone, not just the owners of the tree, but are not allowed to be sold or traded with. This is to stop the temptation for farmers to work the land and trade the produce.

We write this as

$$F_D(\text{trade fruit of tree})$$

We *do* want you to eat the fruit and not to sell them. We do not require in practice to eat the fruit. The Talmudic O_T eat is not enforced. I.e. you have no actual obligation to eat the fruit only not to sell them.

> Leviticus 25:1-7
> God spoke to Moses at Mount Sinai, telling him to speak to the Israelites and say to them: When you come to the land that I am giving you, the land must be given a rest period, a sabbath to God. For **six years you may plant your fields, prune your vineyards, and harvest your crops, but the seventh year is a sabbath of Sabbaths for the land**. It is God's sabbath during which you may not plant your fields, nor prune your vineyards. Do not harvest crops that grow on their own and do not gather the grapes on your unpruned vines, since it is a year of rest for the land. [What grows while] the land is resting may be eaten by you, by your male and female slaves, and by the employees and resident hands who live with you. All the crop shall be eaten by the domestic and wild animals that are in your land.
>
> Leviticus 25:20-22
> And if ye shall say: 'What shall we eat the seventh year?' behold, we may not sow nor gather in our increase'; then I will command My blessing upon you in the sixth year, and it shall bring forth produce for the three years. And ye shall sow the eighth year, and eat of the produce, the old store; until the ninth year, until her produce come in, ye shall eat the old store.

To sharpen and clarify the distinctions between Type 1A and Type 1C, note that during Sukkot, the feast of Tabernacles, we must eat our meals inside the Sukkah, a temporary hut you build in your garden. However if you do eat outside the Sukkah, no punishment is due. It is not clear how to formalise it. Opinions differ, it is either of Type 1A or of Type 1C. The book *Minhat Hinuch* says that if we adopt Type 1A, then if one uses a stolen Sukkah one has not fulfilled his obligation, since he committed a violation in the process, however, if we adopt the view that the Type is 1C, then he has fulfilled his obligation.

Compare with Type 2. For a prohibition of Type 2, of the form $F_T X$, if we violate it and do perform X we get punished! We do not get punished if we violate Type 1A or Type 1C.

Table 2.

$O_T X$	If you do X	If you do $\neg X$
Type 1A, in this case we also have $F_D \neg X$ and consequently $O_D X$	You obeyed the will of God. God rewards you in Heaven	You committed violation. You will have to face the consequences in Heaven.
Type 1B, in this case we do not have $F_D \neg X$.	as above	The incident is not recorded in Heaven
Type 1C, in this case we only have $F_D \neg X$ without having $O_T X$	Your obeyance is not recorded in Heaven	You committed violation. You will have to face the consequences in Heaven.
Comment	If you obey 1A by committing a violation which harms other people, then obligation of type 1A is not fulfilled (you are still considered as having committed violation of 1A) but even under these circumstances an obligation of type 1C is fulfilled. See the book *Minhat Hinuch*.	

Type 2. Full prohibition

The Bible forbids the eating of pork.

$$F_T \text{ (eating pork), and we do not have } O_T(\neg \text{ eat pork}).$$

> Leviticus 11:7-8
> And the pig, because it is parts the hoof and is cloven-footed but does not chew the cud, is unclean to you. You shall not eat any of their flesh, and you shall not touch their carcasses; they are unclean to you.

Type 3. Strong obligation/prohibition

This has the structure

$$O_T \neg X \wedge F_T X$$

An example of this is the Biblical obligation/prohibition about work on the Sabbath (seventh day). We have, for $X =$ doing work, an obligation not to do work and also a prohibition on working. So this is a very strong demand from God!

Another example, if you have a house with accessible roof you must install a railing to the roof to prevent people falling off the roof. This can be interpreted as a typical safety rule. Its status is that of a weak obligation introduced for good practice. If you obey it, you will earn the good will of God. There is also prohibition on being without a railing. So if you do not obey it, there is no punishment. We formalise this by writing

$$O_T \text{ Rail and } F_T \neg \text{ Rail.}$$

To quote the Bible:

> Deuteronomy 22:8
> When you build a new house, you must build a railing around the edge of its flat roof. That way you will not be considered guilty of murder if someone falls from the roof.
>
> Note that in the Sabbath example the Obligation is on lack of action and the prohibition is on action and in the roof example the obligation is on action and the prohibition is on lack of action.

2.3 Contrary to Duties

Type CTD I. Obligation with positive contrary to duty

You should not steal and if you steal you should return what is stolen. We can write:

1. $F_T S$

2. $(S \to O_T R)$

or maybe the dyadic formalisation:

2a. $S \to O_T(R/S)$

Type CTD II. Temporal chain of CTDs

This example is from the Bible.

1. You should not rape a woman.
2. If you do rape a woman you must marry her.[49]
3. If you marry the woman you raped you can never divorce her.

We write this as

$$F_T R \wedge (R \to O_T(M/R)) \wedge (R \wedge M \to F_T(D/R \wedge M))$$

To quote the Bible:

> Deuteronomy 22:28-29
> If a man happens to meet a virgin who is not pledged to be married *and rapes her* and they are discovered, *he shall pay the girl's father fifty shekels of silver. He must marry the girl, for he has violated her. He can never divorce her as long as he lives.*

Type CTD III. Fine tuning required

Let us give some more examples of Contrary to Duties from the Talmud. These examples require further fine tuning and their delicate formalisation is postponed.

1. This is the mainstream example we mentioned before, which we recall here for comparison, that we should not steal but if we do steal we have an obligation to return the stolen property to its rightful owner. (This is a 'repairing' CTD.)
2. We have an obligation to pray three times a day. A morning prayer, an afternoon prayer and an evening prayer. The time for the afternoon prayer is from noon to sunset. The evening prayer should be done after sunset but before sunrise. The rules governing this are as follows:
   (a) It is obligatory to pray the afternoon prayer between noon and sunset.
   (b) If one was not able, due to circumstances beyond his control, to offer the afternoon prayer before sunset one can still fulfill the obligation by offering the afternoon prayer 13 minutes after sunset. (This is called 'make up'.)

[49] Assuming she is not married. If she is married, the guy is in really serious trouble! If she is not married but does not want to marry the guy, he has to pay compensation only.

(c) If time has passed and no afternoon prayer was offered then one can offer the evening prayer twice, to make up for the afternoon prayer.

3. Another example is the Yevama example. If a woman becomes a widow without children and her deceased husband has an unmarried brother, then the brother has a duty to marry the widow to continue the family line. If the brother does not want to do that, he has the duty to give the widow a special ‘divorce’ document to enable her to be free to marry. (This is a ‘way out’, it is not a CTD or a ‘making up’.)

4. A fourth example is the reading of the Book of Esther during the Purim festival. The obligation is to read it standing, not sitting. This is the a priori obligation. But if the reading was done sitting down, it does a posteriori discharge the reader from his obligation. The Talmud makes a distinction between our obligations before the event (‘*Lechatchila*’) and what is required after the event (‘*Bede'eved*’).

5. There are many more cases, for example where the same action violates several prohibitions and obligations, some of them contradictory. These are solved in practice (see Section 3).

REMARK 3. The prayer examples and the Yevama example, are very interesting. They hint to a type of contrary to duties which fulfil the original obligation and are not necessarily just secondary obligations, which kick into action when the original obligation is violated. The CTD can actually cancel the original violation. It is not a disjunction. We do not have the disjunctive option of either reading the Book of Esther standing or sitting. We should a priori try to read it standing but if we read it sitting the original obligation to read it sitting is discharged. In comparison, if I steal a book and then return it, I am still in violation of the ‘do not steal’ obligation. The difference is whether the obligation relates to the process or to the resulting state (after the process).

Let us further remark about the logic involved is the nature of the CTDs in the Talmud. There should be more emphasis on resolving conflicting obligations and prohibitions. The system is built for people to use and live by day-by-day. So the most important feature of the logic is to resolve conflicting obligations and prohibitions arising from a multitude of CTD all triggered by past actions. For this again we need a labelled system. Let us give a modern example to show what we mean and thus realise that ordinary deontic logic has not fully addressed such problems.

EXAMPLE 4. Suppose our starting point is that we have the following:

1. There should be no fence.

2. There should be no dog
3. If there is a dog there should be a fence
4. If there is a fence it should be white
5. If there is a dog and a fence it should be high
6. If there is a fence and it is not white it should be low

Some stubborn rebellious landlord does the following sequence of actions

(s1) get a dog

(s2) build a fence

(s3) paint the fence green.

He now decides to be a good boy and asks for our recommendation of what to do about his violation. Should he at least get a builder and modify the fence and make the fence high or low? How do we proceed?

First let us label his actions by the violations he performed, and ask at each stage what our recommendation would have been. Then we ask if there is a simple case of reverse actions (e.g. get rid of the dog) which will restore consistency. Then we decide what to recommend.

So this is a special case of controlled revision see [61; 62].[50]

In anticipation of formulating a formal system for Talmudic logic, let us say that we probably need to extend **SDL** by allowing labelled formula and include a revision operator $*A$ (A revised) in the object language.

The reader should be aware that the Talmudic way of resolving conflict is different and new to the traditional methods. So there is novelty in that.

Note that Talmudic CTDs have special features as discussed in Remark 3. We can write OX and the contrary to duty saying that if in practice you

[50]The following is the labelled history of actions violations. "+" means obeyance, "−" means violation.

(s1) label $[(-b)]$

(s2) label $[(-b), (+c), (-a)]$

(s3) label $[(-b), (+c), (-a), (-d)]$.

If he makes the fence low we will get also $(+f)$ and $(-e)$, and if he makes the fence high we will also have $(-f)$ and $(+e)$.

On the basis of the above history of labels we make a decision.

Controlled revision applies when we start with a theory Δ_0 and have a series of inputs $A_1, A_2, A_3 \ldots$. At stage n we have Δ_n, and when we revise to accommodate A_{n+1} we must remember the entire history of revisions and revise accordingly.

So, for example, if $\Delta_0 = \{A, A \to B\}$ and we get $\neg B$, we revise and get $\Delta_1 = \{\neg B, A \to B\}$. If we now get input B, we ordinarily may revise and get $\Delta_2 = \{B, A \to B\}$. But in controlled revision we remember the history, so we know that we took out A and hence we bring it back and revise to $\Delta_2^{\text{controlled}} = \{A, A \to B\}$.

have done X' then we consider OX as having been obeyed. So we can write OX and $\neg X \to OX'$ and if X' then there is no violation of OX.

EXAMPLE 5. To give you a glimpse of Talmudic style conflict resolution consider the following two obligations

1. you should always be seen wearing a black suit at official receptions
2. you must always wear a dark blue dinner suit at evening formal dinners.

You get a conflict when invited to an evening do with Her Majesty The Queen. What to wear black or dark blue? Modern non monotonic logic will say rule 2 is more specific, so it has priority. Talmudic reasoning also accepts that the more specific norm may have priority, but in this case we have another simple option: Talmudic style conflict resolution will say that in the evening in electric light dark blue looks black. So there is no conflict! Note that this is not a logical solution but a practical one.

Note that we can give a practical solution to Example 4 by recommending a low fence. Since the fence is painted green, it blends with the grass and plants and can be considered as not violating the obligation that there should be no fence, but only in this case!

2.4 Discussion: Talmudic deontic logic?

The perceptive reader might wonder what kind of (Talmudic) logic we have here. We possibly have the ordinary deontic logic **SDL** for the operators $\{O_D, F_D\}$ and we have two new completely unrelated Talmudic modalities O_T and F_T. We also have lots of examples for them. So where is the logic?

Our answer to this is twofold:

1. Consider a modal, possibly intuitionistic, logic with three separate **KD** modalities generated by O_T, F_T and O_D and study the correct axioms governing them.
2. We can equivalently regard $O_T X$ and $F_T X$ not as modalities but as labels. So each wff X will have several possible labels.

   (a) neither $O_T X$ nor $F_T \neg X$

   (b) $O_T X$ only

   (c) $F_T \neg X$ only

   (d) both $O_T X$ and $F_T \neg X$

   So the logic would be standard deontic logic applied to labelled formulas. This approach also goes well with the fact that O_T and F_T obligation and prohibition carry reward or punishment for obeyance and violations respectively. So the labels can be used to indicate that information as well. Modal systems with labels exist in the literature primarily as Gentzen or tableaux systems and there is work by D. Gabbay and others in this direction [54; 60]. So it should not be difficult to tailor a suitable Talmudic labelled variant of **SDL**. Our guess is that the system should also be intuitionistic, as we have already mentioned earlier.

We now address the problem of formulating an axiom system and semantics for Talmudic deontic logic.

Our first task is to understand the data better. We say that various prohibitions and obligations come in the Talmud from a divine Biblical source (annotated by O_T and F_T). We also know that we may have conflicting obligations and prohibitions emanating possibly directly from the divine source or because of a history of violations and the triggering of contrary to duties. We need to understand how to move from the T operators to the D operators. Once we understand how the Talmud does this, we can construct a logic.

So, before we offer a logic, we need to record and understand this body of data, and the way the Talmud handles conflict resolution.

To focus our thoughts, let us consider an artificial, but familiar example. (Compare with Example 4 and Footnote 50.)

EXAMPLE 6.

1. There should not be a fence.
2. There should not be a dog
3. If there is a dog there should be a fence
4. There is a dog

Let us pretend that the above are Talmudic obligations and prohibitions, given to us as follows:

1. F_T (fence)
2. F_T (dog)
3. dog $\to O_T$ (fence)
4. dog.

Notice that in whatever Talmudic logic we are going to formulate, we may not get the traditional paradox because although we can derive O_T (fence) $\wedge$ F_T (fence), we have two different independent operators involved.

The Talmud is practical and so it needs to tell us what to do in this case.

Imagine a man comes to the Rabbi with a dog and says "Advise me; fence or no fence?".

The data is

1. F_T (fence) — a direct prohibition.
2. O_T (fence) — an obligation arising after a violation of an F_T prohibition.

A decision needs to be made.

We use O_D to indicate practical obligations, the ones which are the results of the Talmudic rules for conflict resolution which enable us to move from the T operators to the D operators and thus equip us with the tools of making day to day practical decisions. This answer is independent of A, as there is a general rule that $O_T(A)$ is stronger than $F_T(A)$ for any A.

Let us say the Rabbi tells our man to do a fence, (i.e. O_D (fence)), then we have the following decision table, Table 3.

Table 3.

	$F_T A$
$O_T A$	$O_D A$

Here we ignored the fact that $O_T A$ is a result of a CTD violation. Table 3 says simply that if there is a conflict between $F_T A$ and $O_T A$, then you do A, i.e. $O_D A$ is the answer.

Let us make this example more complicated. Let us add another obligation to maintain a well kept garden and the contrary to duty that if we do not do so, then we have to have a fence.

So we add to our example

5. O_T (well kept garden)
6. $\neg$ well kept garden $\rightarrow O_T$ (fence)
7. $\neg$ well kept garden.

Now the conflict is between two cases of O_T (fence) (later in the formal system we shall add an index to the T operators to enable us to represent several different uses of them) and one case of F_T (fence).

It is important to note that the Talmud regards contrary to duties as context dependent obligations and prohibitions and gives them equal standing as any other obligations and prohibitions. This is compatible with the dyadic view of contrary to duties, where we write $O_T(X/Y)$ and $F_T(X/Y)$. The Talmud even numbered all existing obligations and prohibitions; there are 248 generic obligations and 365 generic prohibitions, some of them are

CTDs and some are not. So really we should write $O_{(1,T)}, \ldots, O_{(248,T)}$ and $F_{(1,T)}, \ldots, F_{(365,T)}$. So any single specific situation may potentially fall under 613 conflicting obligations and prohibitions. In our example for the single question of having a fence the number is 3.[51]

In fact, we shall find that the correct modelling of all obligations and prohibitions in the Talmud is by dyadic operators, $O_T(X/Y)$ and $F_T(X/Y)$, where Y is a context. The contrary to duties are cases where there is a violation and therefore the context changes to include the violation details. This also explains why the Talmud does not pay special attention to contrary to duties. All obligations and prohibitions are context dependent anyway!

The next step for us is to document a full conflict resolution table as practiced by the Talmud. If we take the dyadic view of the Obligations and prohibitions

$$O_{(1,T)}(X_1/Y_1), \ldots, O_{(248,T)}(X_{248}/Y_{248})$$

and

$$F_{(1,T)}(U_1/V_1), \ldots, F_{(365,T)}(U_{365}/V_{365}),$$

we get conflict between several obligations and prohibitions in contexts Z common to several Y_i and V_j. It is in such cases that the Talmud offers conflict resolution. The interesting aspect of the Talmudic conflict resolution is that it does not depend on the context Z or on how many previous violations were committed in the way to the context Z but it depends purely on the form $O_T(A)$, or $O_T(\neg A)$ or $F_T(A)$ or $F_T(\neg A)$ of the conflicting participants where A denotes the action discussed.

A close inspection of the Talmud reveals that the underlying logic should be intuitionistic based possibly on decided atomic facts. So doing this for our dog example, we have, (we are simplifying and not counting multiple instances of O_T and F_T, for example $O_{(i,T)}$, as above.).

Fact in question: $A =$ dog.
We want to decide

$$\text{fence} \vee \neg \text{ fence}$$

In our dog example, the Rabbi can tell the man one of the following options:

1. O_D (fence): you must have a fence
2. O_D ($\neg$ fence): you must not have a fence
3. $\neg\neg O_D$ (fence): I can only recommend that the decent thing to do is not to have a fence
4. $\neg\neg O_D(\neg$ fence): I recommend the decent thing to do is to have a fence

[51] For a discussion of why it is necessary to use 613 labelled modalities, see Section 5.2.

5. $\varnothing$: no decision, no comment, do whatever you want.

Considering the general case, there are twelve options and these options are listed in Table 4. This table is intuitionistic. Note that A itself, being a fact, is classical, i.e. $\neg\neg A \equiv A$ holds. Note that A is the action of having a fence.

Table 4. List of 12 possibilities for Talmudic obligations and or prohibitions for A

1.	$F_T(A)$	A is prohibited
2.	$\neg F_T(A)$	A is not prohibited
3.	$F_T(\neg A)$	Not doing A is prohibited
4.	$\neg F_T(\neg A)$	There is no prohibition on not doing A
5.	$\neg\neg F_T(A)$	A is not prohibited but the right mode of behaviour is not to do A, i.e. weak prohibition of A
6.	$\neg\neg F_T(\neg A)$	$\neg A$ is not prohibited but the right mode of behaviour is not to do A, i.e. weak prohibition of $\neg A$
7.	$O_T(A)$	A is obligatory
8.	$\neg O_T(A)$	A is not obligatory
9.	$O_T(\neg A)$	$\neg A$ is obligatory
10.	$\neg O_T(\neg A)$	$\neg A$ is not obligatory
11.	$\neg\neg O_T(A)$	A is not obligatory but good mode of behaviour is to do A
12.	$\neg\neg O_T(\neg A)$	$\neg A$ is not obligatory but good mode of behaviour is not to do A

We now form two 12×12 tables indicating how to resolve conflicts between the elements of Table 4. Table 5 indicates the conflict resolution in terms of the T operators, and Table 6 indicates, on the basis of Table 5, what should be done in practice. We use the intuitionistic operator $O_D A$ in Table 6. Thus Table 5 and Table 6 together give the Talmudic conflict resolution strategy. For example entry $(1,3)$ of Table 5 is $F_T(A)$ pitted against $F_T(\neg A)$, and by $P6$ (b) below, $F_T(A)$ wins, and so in Table 5 we put "1" in box $(1,3)$, namely we put "$F_T(A)$" in box $(1,3)$. Then in Table 6 we put in $O_D(\neg A)$, to indicate what we do in practice.

We now describe the Talmudic principles behind the construction of Table 5 and Table 6.

(P0): The entries in Table 6 uses "$O_D(A)$", "$O_D(\neg A)$", "$\neg O_D(A)$", "$\neg O_D(\neg A)$", "$\neg\neg O_D(A)$" and "$\neg\neg O_D(\neg A)$". When we write "$\neg O_D(A), \neg O_D(\neg A)$" as a single entry , we mean that no decision

Table 5. Conflict resolution table for the T operators of Table 4

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	1	1	1	1	1,5	1	7	1	1,9	1	1	1,12
1		2	3	2,4	5	6	7	2,8	9	2,10	11	12
3			3	3	3	36	3,7	3	9	3,10	3,11	3
4				4	5	6	7	4,8	9	4,10	11	12
5					5	5	7	5	5,9	5	11	5,12
6						6	6,7	6	9	6	6,11	12
7							7	7	9	7	7,11	7
8								8	9	8,10	11	12
9									9	9	9	9,12
10										10	11	12
11											11	12
12												12

Table 6. Conflict resolution table for T operators of Table 4 and their D operator result

	1	2	3	4	5	6
1.	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$
2.		$\neg O_D(A), \neg O_D(\neg A)$	$O_D(A)$	$\neg O_D(A), \neg O_D(\neg A)$	$\neg\neg O_D(\neg A)$	$\neg\neg O_D(A)$
3.			$O_D(A)$	$O_D(A)$	$O_D(\neg A)$	$O_D(A)$
4.				$\neg O_D(A), \neg O_D(\neg A)$	$\neg\neg O_D(\neg A)$	$\neg\neg O_D(A)$
5.					$\neg\neg O_D(\neg A)$	$\neg\neg O_D(\neg A)$
6.						$\neg\neg O_D(A)$
7.						
8.						
9.						
10.						
11.						
12.						

	7	8	9	10	11	12
1.	$O_D(A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$
2.	$O_D(A)$	$\neg O_D(A), \neg O_D(\neg A)$	$O_D(\neg A)$	$\neg O_D(A), \neg O_D(\neg A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
3.	$O_D(A)$	$O_D(A)$	$O_D(\neg A)$	$O_D(A)$	$O_D(A)$	$O_D(A)$
4.	$O_D(A)$	$\neg O_D(A), \neg O_D(\neg A)$	$O_D(\neg A)$	$\neg O_D(A), \neg O_D(\neg A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
5.	$O_D(A)$	$\neg\neg O_D(\neg A)$	$O_D(\neg A)$	$\neg\neg O_D(\neg A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
6.	$O_D(A)$	$\neg\neg O_D(A)$	$O_D(\neg A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
7.	$O_D(A)$	$O_D(A)$	$O_D(\neg A)$	$O_D(A)$	$O_D(A)$	$O_D(A)$
8.		$\neg O_D(A), \neg O_D(\neg A)$	$O_D(\neg A)$	$\neg O_D(A), \neg O_D(\neg A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
9.			$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$	$O_D(\neg A)$
10.				$\neg O_D(A), \neg O_D(\neg A)$	$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
11.					$\neg\neg O_D(A)$	$\neg\neg O_D(\neg A)$
12.						$\neg\neg O_D(\neg A)$

is made, no comment . The table is symmetrical, so we give only the entries above the diagonal.

(P1): Items 1, 3, 7 and 9 are ordinary Biblical norms.

(P2): Items 5, 6, 11 and 12 are not demands (norms) from God, but it would please God if we adopt them. In many practical cases the Rabbis and the courts force people (legislate) to adopt them.[52]

(P3): Items 2, 4, 8 and 10 mean that there is no relevant normative obligation or prohibition regarding A.

(P4): In any conflict between items in (P1) and items in (P2) and (P3), (P1) should win.

(P5): In any conflict between items in (P3) with items in (P1) or (P2), (P3) should *not* win.

(P6): In conflict between items in (P1) itself, the following are the rules:

(a) O_T is stronger than F_T, i.e. we always prefer positive norm, so O_T wins and therefore Table 5 gives O_T and Table 6 gives O_D see, for example, entry $(1, 7)$. The result is therefore O_D.

(b) In any conflict between $O_T A$ and $O_T \neg A$ or in any conflict between $F_T A$ and $F_T \neg A$, we always prefer to do nothing, hence $O_T(\neg A)$ and $F_T(A)$ respectively win, and therefore the entry in Table 6 is $O_D \neg A$.

Note that we need to use a mechanism to determine for each case A and $\neg A$ which one is the action and which none is the negation of action. So for example if we have O_T(sleep) and O_T(be awake) our mechanism needs to determine which one we consider action and hence call it A and which one the lack of action and call it $\neg A$. We assume that it is always clear to us which option between A and $\neg A$ is the action

[52] To give an example, suppose I find a lost item in the street, say a handkerchief. There are two possibilities to consider.

1. The owner does not bother to come back looking for it.
2. The owner will not give up and come back for it (monogrammed handkerchief).

Legally in case 1 my obligation to seek the owner does not exist since the owner has given up. In comparison in the second case I must pick up the handkerchief and find the owner or give it to the police.

However, even in the first case, it is recommended and even legislated that I try and find the owner (e.g. give it to the police), even though the owner has abandoned the handkerchief , i. e. there is no O_T obligation to return the handkerchief, but nevertheless the Talmud recommends that I return the handkerchief. Our notation for this is $\neg\neg O_T$.

(P7): When there is conflict inside group (P2), there is no clear cut rule. It is reasonable that, since O_T is stronger than F_T, we also should have that $\neg\neg O_T$ is stronger than $\neg\neg F_T$.

Similarly it is reasonable to resolve conflict between $\neg\neg O_T A$ and $\neg\neg O_T \neg A$ or $\neg\neg F_T A$ and $\neg\neg F_T \neg A$ by choosing the "lack of action" option, i.e. $\neg\neg O_T \neg A$.

(P8): Conflicts among (P3) are meaningless. All is open and possible.

2.5 Intuitionistic standard deontic logic

In Section 3 we analysed the conflict resolution method in the Talmud for the operators O_T and F_T and how they relate to O_D.

Our conclusion from Example 6 is that we need 248 different $O_T(X/Y)$ operators and 365 different $F_T(U/V)$ operators. This should not alarm the deontic logician because in the dyadic approach, we have an infinite number of operators $O_A(X)$, one for each wff A. The difference between dyadic deontic logic and the Talmudic ones is that the Talmudic operators are generic, and apply to an ever growing open texture context situations Y and V. Let us simplify and begin by giving an Intuitionistic Standard Deontic Logic with one unary (not dyadic) O_T and one unary F_T, just to be able to compare ordinary classical **SDL** with the intuitionistic version of it. Remember also that once we fix the context Z of conflict, the dyadic operators become monadic (relative to Z), so our monadic logical machinery is applicable anyway.

We are now ready to address axiom systems and semantics.

Our strategy is to first give the operators O_T, F_T and O_D (we don't seem to need F_D!) suitable semantics and see whether Table 6 can be derived from some semantic principles. In formal logic this means that the system for O_D would be a nonmonotonic intuitionistic modal logic derived in some way from a monotonic intuitionistic modal logic for $\{O_T, F_T\}$. Note that this is a sound policy, as in the area of nonmonotonic logic, the classification of nonmonotonic logics is done in terms of variations on an underlying monotonic base logic.

The reader should beware that our system for O_D must be derived nonmonotonically from the system for $\{O_T, F_T\}$. If we look at Table 6, we see that for example, the conflict between $O_T A$ and $F_T A$ is resolved as $O_D A$. We must not be tempted and to simply write the axiom

$$O_T A \wedge F_T A \to O_D A$$

and similarly write more rules for each box in the table.

This would lead to contradictions because the logic we need is nonmonotonic.

To see this note that if we have $O_T(A)$ alone we get $O_D(A)$, and if we have $F_T(A)$ alone we get $O_D(\neg A)$. However if we have both of them we get $O_D(A)$. A monotonic logic would give us pragmatic oddity, $O_D(A) \wedge O_D(\neg A)$. For this reason we use Remark 10 below to define the semantics for O_D. Semantically when we have A we have the set of all worlds where A holds and hence we know the totality of all T operators it satisfies and hence we can make a decision based on all the data about A.

We begin with a semantical presentation of an intuitionistic modal logic with O^i_T and F^j_T $i = 1, \ldots, 248$ and $j = 1, \ldots, 365$. For simplicity, since all the operators are independent, we shall deal with a system with only two O_T and two F_T operators, and the usual intuitionistic connectives $\neg, \wedge, \vee, \rightarrow, \top, \bot$.

So we have O^1_T, O^2_T and F^1_T and F^2_T.

We need two of each because of entries in Table 6. For example entry $(1, 3) = (F_T A, F_T \neg A)$ really means $(F^1_T A, F^2_T \neg A)$, where the two prohibitions F^1_T and F^2_T contradict. Similarly, entry $(7, 9) = (O^1_T A, O^2_T \neg A)$.

DEFINITION 7. A model has the form $\mathbf{m} = (S, \leq, I^i_O, I^i_F, a, h), i = 1, 2$ where $(S, \leq, a)$ is a partially ordered set of worlds and $a \in S$ is the actual world and I^i_O, and I^i_F $i = 1, 2$ are functions associating a nonempty set of worlds with each $t \in S$.

We have

$$t \leq s \Rightarrow I(t) \supseteq I(s)$$

for each I.

For each atomic $q, h(q) \subseteq S$ is an assignment to the atoms. We have

$$t \leq s \Rightarrow t \in h(q) \rightarrow s \in h(q).$$

Satisfaction is defined as follows:

1. $t \vDash q$ iff $t \in h(q)$
2. $t \vDash A \wedge B$ iff $t \vDash A$ and $t \vDash B$
3. $t \vDash A \vee B$ iff $t \vDash A$ or $t \vDash B$
4. $t \vDash A \rightarrow B$ iff for all $t \leq s, s \vDash A$ implies $s \vDash B$
5. $t \vDash \neg A$ iff for all $s, t \leq s \Rightarrow s \nvDash A$
6. $t \vDash F^i_T A$ iff for all $s \in I^i_F(t), s \nvDash A$
7. $t \vDash O^i_T A$ iff for all $s \in I^i_O(t), s \vDash A$
8. $t \vDash \top$ and $t \nvDash \bot$
9. $\mathbf{m} \vDash A$ if $a \vDash A$.

DEFINITION 8. We offer the following axiom system **ISDL**

1. Axioms and rules for intuitionistic logic for $\{\neg, \wedge, \vee, \rightarrow, \top, \bot\}$.
2. $O^i_T A \wedge O^i_T(A \rightarrow B) \rightarrow O^i_T B$
3. $F^i_T(A \vee B) \leftrightarrow F^i_T A \wedge F^i_T B$
4. $\vdash O^i_T \top, \vdash F^i_T \bot, \vdash \neg O^i_T \bot$ and $\vdash F^i_T \top$
5. $\frac{\vdash A \rightarrow B}{\vdash O^i_T A \rightarrow O^i_T B}$
6. $\frac{\vdash A \rightarrow B}{\vdash F^i_T B \rightarrow F^i_T A}$

THEOREM 9. **ISDL** *is complete for the proposed semantics.*

Proof.

1. Let S be the set of all complete consistent theories (Δ, Θ). Completeness means that for each A, $A \in \Theta$ or $A \in \Delta$ and consistency means that for no $A_i \in \Delta, B_j \in \Theta$ do we have $\vdash \bigwedge A_i \rightarrow \bigvee B_j$. We know that every consistent theory (Δ, Θ) can be extended to a complete and consistent theory (Δ', Θ'), with $\Delta \subseteq \Delta', \Theta \subseteq \Theta'$.
2. Define $(\Delta_1, \Theta_1) \leq (\Delta_2, \Theta_2)$ iff $\Delta_1 \subseteq \Delta_2$
3. Define $(\Delta', \Theta') \in I^i_O((\Delta, \Theta))$ iff for all $O^i_T X \in \Delta$ we have $X \in \Delta'$.
4. Define $(\Delta', \Theta') \in I^i_F((\Delta, \Theta))$ iff for all $F^i_T X \in \Delta$ we have $X \in \Theta')$
5. Lemma: If $O^i_T X \in \Theta$ then for some $(\Delta', \Theta') \in I^i_O((\Delta, \Theta))$ we have $X \in \Theta'$.

   Proof. Consider $(\{Y | O^i_T Y \in \Delta\}, \{X\})$. We claim this theory is consistent. Otherwise for some $O^i_T Y_i \in \Delta$, we have

   $$\vdash \bigwedge Y_i \rightarrow X$$

   Hence

   $$\vdash \bigwedge O^i_T Y_i \rightarrow O^i_T X$$

   and we get $O^i_T X \in \Delta$, a contradiction.

   Extend the above set to a complete consistent theory (Δ_1, Θ_1) and this theory is what we need.

6. If $F^i_T X \in \Theta$, then for some $(\Delta_1, \Theta_1) \in I^i_F(\Delta, \Theta)$ we have $X \in \Delta_1$.

   Proof. Consider
   $$(\{X\}, \{Y | F^i_T Y \in \Delta\}.$$
   We claim this set is consistent. Otherwise for some $F^i_T Y_i \in \Delta$, we have
   $$\vdash X \to \bigvee Y_i.$$
   Hence $\vdash F^i_T \bigvee Y_i \to F^i_T X$ and hence
   $$\vdash \bigwedge F^i_T Y_i \to F^i_T X$$
   and hence $F_T X \in \Delta$ a contradiction.

   Extend the above set to a complete consistent theory (Δ_1, Θ_1) as needed.

7. **Lemma**
   Let $h(q) = \{(\Delta, \Theta) | q \in \Delta\}$. Then in the model $(S, \leq I_O, I_F, h)$ we have for each A $(\Delta, \Theta) \vDash A$ iff $A \in \Delta$.
   Proof.
   By induction on A

■

REMARK 10. We now add O_D to our model. Semantically we define a neighbourhood $\mathcal{N}(t)$ for each $t \in S$ and let $t \vDash O_D A$ iff $\{s | s \vDash A\} \in \mathcal{N}(t)$.

We define $\mathcal{N}(t)$ according to Table 6.

Let
$$+A = \{s | s \vDash A\}$$
$$-A = \{s | s \nvDash A\}.$$
We compare the sets $\pm A$ with the sets $I^i_O(t)$ and $I^i_F(t)$, $i = 1, 2$ and decide according to Table 6 whether to admit $+A$ into $\mathcal{N}(t)$.

Note that because O_T and F_T are intuitionistic, the intuitionistic condition below is fulfilled for O_D as well, namely
$$t \leq s \to \mathcal{N}(t) \subseteq \mathcal{N}(s)$$
Therefore
$$t \vDash O_D A \wedge t \leq s \Rightarrow s \vDash O_D A \text{ holds.}$$

Note that the definition of the logic for $\{O_T, F_T, O_D\}$ is semantic. We take a model of $\{O_T, F_T\}$ and add to it $\mathcal{N}$ and make it a model of O_D.

Before we continue we need to make an important remark. Suppose a person embarks on a sequence of actions, just doing whatever he wants for a while. He may find himself in a state where a sequence of obligations and prohibitions has been triggered and some of these may be conflicting.

In fact an obligation to do some A may have been triggered several times in different contexts for different reasons and similarly the obligation to do $\neg A$, as well as the prohibition to do A and the prohibition to do $\neg A$.

The big question now is what to do? In other words do we have $O_D A$ or do we have $O_D \neg A$?

We propose to use Table 6 for this purpose.

We need to check the coherence of Table 6, namely that we get a clear answer for each subset $\pm A$ whether it should be a member of $\mathcal{N}(t)$. This belief follows from conditions (P1)–(P8), especially (P4)–(P8) which clearly set out conflict resolution rules once we understand the semantic properties of O_D we can try to axiomatise it.

If it turns out that we do not have a unique answer in each case we need an expanded new table resolving conflicts between 4 operators at a time and not just two at a time. In fact Remark 11 below shows that the table does give unique answers. Before we systematically do all the cases, let us explain the method by giving two examples.

Take for example the triple set $\{1 : O_{1,T}(A), 2 : O_{2,T}(\neg A), 3 : F_{1,T}(A)\}$, and let us combine them in different orders and see whether we get the same outcome.

Case 1: 2 and 3 give 2 as a winner , see (P6), and now that we have 2, we carry on; 2 and 1 give 2 again by (P6) and the resolution is $4 : O_D(\neg A)$
Case 2: 1 and 2 give 2 and 2 and 3 give 2 and the resolution is again 4
Case 3: 1 and 3 give 1 and 1 and 2 give 2 and the resolution is again 4.
This is because according to Table 6, 2 is the strongest.

So the answer is 4 no matter at what order we combine them.

Indeed this always true that we get a single answer for any group of 3 or 4 items. The Talmud however, does not always agree with the table. As we shall see in Remark 11 below, the table is only a very good approximation. Let us look at another example: Consider the triple $\{O_{1,T}(A), F_{1,T}(\neg A), O_{2,T}(\neg A)\}$, Table 6 will give the clear cut result $O_D(\neg A)$. However the Talmud in this case decrees that the combined power of $O_{1,T}(A), F_{1,T}(\neg A)$ together (which yields according to the usual priorities $O_T(A)$), is stronger than $O_{2,T}(\neg A)$, and so the result should be $O_D(A)$. This clearly shows that Table 6 is only an approximation of the way the Talmud combines obligations.

REMARK 11 (Coherence of Tables 5 and 6). The more complex situations for Table 5 are the cases of three conliciting prohibitions and obligations. We divide the cases into four classes:

Class 1: One O_T and two F_Ts.
We have two subcases. First we take two F_Ts and compare with O_T or first we take $\{O_T, F_T\}$ and compare with the second F_T.

Case 1.1: $F_{1,T}(A), F_{2,T}(A)$ and $\{O_T(A)\}$.

In this case O_T is stronger according to entry $(1, 7)$ of Table 5.

Although our Table 5 gives a clear cut answer for this triplet case, the Talmud contains a discussion about whether the answer is acceptable. Some Talmudic scholars express the opinion that two $F_T(A)$ can aggregate and be stronger than one $O_T(A)$. Other scholars accept the answer of Table 5 as the correct one.

So to sum up: The table yields $O_T(A)$ as the result for the triplet and thus records $O_D(A)$ in entry (1,,7). In contrast, the opinion which aggregate would like to have $F_T(A)$ as the winner for this case and would therefore recommend $O_D(\neg A)$. But to aggregate we need a new (three dimensional) table, dealing with triplets.

Case 1.2 $F_{1,T}(\neg A), F_{2,T}(A)$, and $O_T(A)$.

Case 1.3 $F_{1,T}(A), F_{2,T}(\neg A)$, and $O_T(\neg A)$

In both cases, by (P6) we get the same result, namely that O_T wins. In case 1.2 $O_T(A)$ wins, and in Case 1.3 $O_T(\neg A)$ wins.

Class 2: Two O_Ts and one F_T.
Here we have $O_{1,T}, O_{2,T}$ pitted against F_T. This case is discussed explicitly in the Talmud, and there is an explicit ruling in the Talmud that $O_{1,T}$ is the winner.

Indeed Table 5 and (P6) give the same result.

Case 2.1: $O_{1,T}(\neg A), O_{2,T}(A)$, and $F_T(A)$.
Here the answer is immediate, $O_{1,T}(\neg A)$ is the winner. So Table 6 gives the answer $O_D(\neg A)$.

Case 2.2: $O_{1,T}(A), O_{2,T}(\neg A)$, and $F_T(\neg A)$.
This is a clear cut case. The answer is $O_D(\neg A)$. However the intermediate calculations using (P6) have one interesting feature.

When we combine $O_{1,T}(A)$ with $F_T(\neg A)$, we notice that they are not in conflict, but they agree. So (P6) has nothing to say about this. We do need however, a formal answer for the result. Is $O_{1,T}(A)$ the formal winner, or is $F_T(\neg A)$ the formal winner?

Fortunately, when pitted against $O_{2,T}(\neg A)$ we get that $O_{2,T}(\neg A)$ is the winner in either case and so the final answer is $O_D(\neg A)$.

So the table gives us a unique answer $O_D(\neg A)$, however, the Talmud does not accept this result. The Talmud rules that $O_{1,T}(A)$ must win.

This ruling means the Talmud is doing some aggregation and thus gets a different result from (P6) and from Table 5.

We can say that the Talmud aggregates $O_{1,T}(A)$ and $F_T(\neg A)$ which agree and reinforce each other into something stronger, call it $FO_T(A)$, and this wins against the third $O_{2,T}(\neg A)$. Thus $O_D(\neg A)$ is the answer the Talmud wants, contrary to Table 5.

We can formally add FO_T as a new operator and extend Table 4 with 6 additional options:

13. $FO_T(A)$
14. $\neg FO_T(A)$
15. $FO_T(\neg A)$
16. $\neg FO_T(\neg A)$
17. $\neg\neg FO_T(A)$
18. $\neg\neg FO_T(\neg A)$

We can now write a new Table 5, containing 18×18 entries.

However, the Talmud is not clear about some of the entries of such a new table. For example, we do not know the Talmud's view of the strength of $FO_T(\neg A)$.

So we leave Tables 4, 5 and 6 as they are and note that mathematically Table 5 is coherent and as far as our logic is concerned, we are formally OK.

Class 3: Three O_Ts.
Here there is a conflict between two $O_T(X)$ and one $O_T(\neg X)$. We distinguish two subcases.

Case 3.1: $O_{1,T}(A), O_{2,T}(A)$, and $O_{3,T}(\neg A)$.
In this case $O_{3,T}(\neg A)$ is always the winner according to Table 5. The question to ask is do we want to aggregate $O_{1,T}(A)$ and $O_{2,T}(A)$ and make their combined force stronger than $O_{3,T}(\neg A)$?

Indeed, some Talmudic scholars adopt this view, and liken our case to that of Case 1.2, where contrary to Table 5, there is the view of two F_Ts being stronger (when combined) than one O_T.

Case 3.2: $O_{1,T}(\neg A), O_{2,T}(\neg A)$, and $O_{3,T}(A)$.
In this case it is clear that $O_T(\neg A)$ wins according to all combinations using Table 6 and (P6).

Class 4: Three F_Ts.
This case is completely parallel to Class 3 with similar results.

Case 4.1: $F_{1,T}(A), F_{2,T}(A)$ and $F_{3,T}(\neg A)$.
In this case (P6) gives that $F_{3,T}(\neg A)$ is the winner. Here again some scholars might want to aggregate the two $F_T(A)$, with similar discussion to Case 3.1.

Case 4.2: $F_{1,T}(\neg A), F_{2,T}(\neg A)$, and $F_{3,T}(A)$.
In this case it is clear that $F_T(\neg A)$ is the winner.

This concludes our examination of triplets and we verified that Table 5 gives a clear unique answer in each case independent of the order of combination. We noted during our examination that some scholars might want to aggregate, in which case a new table needs to be agreed upon.

Our Table 5 is mathematically coherent for triplets. We now ask: How about sets of four? (P6) and Table 6 is coherent for the Talmud itself does not discuss such cases, and only some later Talmudic scholars raise some examples. We have evidence of discussions of the case of three O_Ts and one F_T.

Case 5: $O_{1,T}(A), O_{2,T}(A), O_{3,T}(\neg A)$, and $F_T(A)$.
Table 6 and (P6) give us the unique answer $O_D(\neg A)$. However, if we start with a choice of triplet and allow for aggregation (which is not according to Table 6, we get two possible answers.

1. If we start with $\{O_{1,T}(A), O_{2,T}(A), O_{3,T}(\neg A)\}$ then $O_{1,T}(A)$ and $O_{2,T}(A)$ aggregate and win and then the winning combined $O_T(A)$ continues to win against $F_T(A)$.

   So the result would be $O_D(A)$.
2. If we start with $\{O_{1,T}(A), O_{3,T}(\neg A), F_T(A)\}$ then $O_{3,T}(\neg A)$ and $F_T(A)$ aggregate and win against $O_{1,T}(A)$ and continue to win further against $O_{2,T}(A)$, and the result would be $O_D(\neg A)$.

   This means that the aggregation point of view is not coherent!

Case 6: $O_{1,T}(\neg A), O_{2,T}(\neg A), O_{3,T}(A)$, and $F_T(\neg A)$.
In this case Table 5 yields the clear cut $O_D(\neg A)$ without any dependence on the order of combination. This case differs from Case 5 in the sense that also different choice of triplets to start with give us the same answer as well, namely $O_D(\neg A)$, even if we consider disagreements on triplets. So the aggregation point of view comes out coherent in this case.

Summary: In the case of 4, we see that those who want to aggregate for the case of 3 are still not coherent for the case of 4. So the only way to remain coherent for all cases is to follow (P6) and Table 5.

REMARK 12. We make an interesting remark about the case of two O_Ts and one F_T. This is Case 2.2, where the Talmud aggregates and disagrees with Table 5 and (P6). However, there is some Talmudic discussion that does not seem to recognise Table 5.

The discussion is about how many violations occur in each case. Consider the case

$$O_{1,T}(\neg A), F_{1,T}(A), \text{ and } O_{2,T}(A).$$

This is Case 2.1. According to Table 6, and a God fearing man should follow $O_D(\neg A)$. Suppose a man decides to do A; we ask how many violations did the man commit? (God punishes for violations!) We might say he violated both $O_{1,T}(\neg A)$ and $F_{1,T}(A)$. This man, however, might argue that he committed only one violation, because if we start with the pair $\{O_{2,T}(A), F_{1,T}(A)\}$ the winner is $O_{2,T}(A)$ and so $F_{1,T}(A)$ being the loser according to the Talmud (as reported in (P6) and Table 5), is out of the picture and hence cannot be violated.

On the other hand, if we start with $\{O_{1,T}(\neg A), O_{2,T}(A)\}$, then the winner is $O_{1,T}(\neg A)$, which agrees with $F_{1,T}(A)$ and so $O_D(\neg A)$ is what our man should have followed and by doing A he violates both $O_{1,T}(\neg A)$ and $F_{1,T}(A)$.

There is a disagreement among scholars with regard to the number of violations in this case, which can be explained by the order in which the triplet is applied.

In contrast to the above, in the case of $\{O_{1,T}(A), F_{1,T}(\neg A), O_{2,T}(\neg A)\}$ (this is Case 2.2), we saw here that according to Table 5 and (P6), $O_T(\neg A)$ wins, but we saw that the Talmud rules that $O_T(A)$ should win, by constructing the aggregated norm $FO(A)$.

Now assume a man does $\neg A$. Here we cannot explain the opinion(of some Talmudic scholars) that $F_{1,T}(\neg A)$ was not violated using an argument concerning the order of combining the norms, because $F_{1,T}(\neg A)$ is aggregated!

We can say that $\neg A$ is a lack of action and claim that one cannot violate in principle any $F_T(\neg A)$, but one can violate $F_T(A)$. However, this does not look convincing. We will ot go into this any further.

Anyway, this remark gives the reader a taste of what is involved in Talmudic argumentation about violations.

REMARK 13. Let us give quick comparisons with the traditional view of obligations and contrary to duties, as described in for example [1,2].

1. Talmudic obligations are generic meta-level and are open texture.

2. The Talmud uses independent obligations and prohibitions.

3. The Talmud regards contrary to duties as obligations/prohibitions arising in some context, and considers them of equal standing with original prohibitions/ obligations. Furthermore, it lists formally 613 such norms, including 248 generic obligations and 365 generic prohibitions,some of them are CTDs and some are not.

4. The Talmud provides rules and tables for conflicts between these 613 norms. It looks only at the form of the norm as in Table 5 or similar

tables and does not look at the content nor consider how the norm was activated by how many violations of how many other norms. Compare this divine approach with [1] which tries to determines logically when an obligation OA can pass on to a contrary to duty context OB. The considerations involve A and B.

Concluding remarks

The remaining subsections will clarify some key points, as promised in the footnotes.

2.6 Reward and punishment

In ordinary Deontic logic and general legal and ethical systems, it is accepted that the difference between obligations and prohibitions manifests itself in the question of whether we are required to take active action or a deliberate lack of action. In comparison in Talmudic thinking the difference between obligations and prohibitions is something different. A biblical obligation is a requirement from a man to better himself and a prohibition is a requirement from a man to make sure he does not decline and deteriorate. The question of whether these requirements are fulfilled and obeyed by a man through his taking action or maintaining lack of action is not important.

The biblical obligation to observe the Sabbath as a holy day is achieved through lack of action (lack of doing any work). This means that the state of a man of not doing work on the Sabbath is a positive state, it makes him a better man, and we are required to achieve this state. In comparison, the biblical prohibition in Leviticus Chapter 19, verse 16 says

> Thou shalt not go up and down as a talebearer among thy people; neither shalt thou stand idly by the blood of thy neighbour: I am the LORD.

The Prohibition

> Neither shalt thou stand against the blood of thy neighbour

is the Good Samaritan Rule, requires us not to stand idle when our neighbour needs our help. It being a prohibition does mean that the Bible views the lack of helping as a negative state (and does not view the act of helping as a positive state, not according to this verse). This means that the Bible views helping and saving your neighbour in need is an elementary requirement, a *natural state for man*, and so obeying this rule does not lead to spiritual betterment but the lack of obeyance of this prohibition can lead to spiritual deterioration.

In comparison, in the case of the Sabbath, the Bible view that *natural elementary state for man* is to go to work on the Sabbath day, and the obligation to refrain from working on the Sabbath day is a requirement intended for the spiritual betterment from the natural state.

Traditional legal and ethical systems do not offer an objective definition of *man's natural state*. So in such systems they have only the distinctions between taking actions and maintaining lack of action. In Talmudic biblical law on the other hand the requirement for the betterment of man's state is given in the Bible as an obligation and the requirement not to deteriorate to a worse state is given as a prohibition.

An example from general legal debate is the problem addressed by Robert Nozick [89; 90] regarding Seduction and Blackmail. What is the difference between the two? On the face of it, in both cases one tries to make his neighbour do something. In the case of seduction, we offer our neighbour a reward for taking the action and in the case of blackmail, we offer him punishment in case of his not taking the action. Here again we see that there is a natural state. Any requirement which is not compatible with it is blackmail and any legitimate requirement which is compatible with it is seduction. If I say to a man that if he does a job for me I shall pay him , this is seduction, because in the natural state I need not pay him. But if I say to the man that if he does not do the job for me I shall beat him up, this is blackmail. The reason for that is that not to be beaten up is an elementary right and natural state and he deserved this right even if he does not do the job for me.

How do we define the dividing line between elementary rights and such that are not? We cannot deal with this here.

The Bible indicates this distinction by the way it presents the obligations and prohibitions. If the requirement is written in the Bible as an obligation(to do an action or to maintain lack of action) then the requirement is compatible with what the Bible regards as a natural state and intends to better it, and if the required is a prohibition, then doing (or lack of doing) what is prohibited is not compatible with the natural state and causes deterioration.

We can now understand why the Bible offers a reward for fulfilling obligations and gives no reward for obeying prohibitions, while also it does not punish for not fulfilling obligations and does punish for disobeying prohibitions. The explanation is that reward is forthcoming to those who better themselves, and if they do not better themselves why punish them? On the other hand if a man deteriorates he should be punished and if he avoids deterioration why should he be rewarded?

We discuss this issue at length in the second part of our book [6].

2.7 Why 613 Talmudic operators?

We now explain why we need so many Talmudic modal operators, 248 for obligations and 365 for prohibitions. Why not have just one operator for obligations and one for prohibitions, as in standard Deontic logic?

The basic claim is that as we go through our daily life we may end up in a situation where several different biblical Obligations and prohibitions apply (coming from different sources in the Bible). For example every seventh year we must let our fields rest and we are not allowed to plough our fields. A man may plough his field on the Sabbath on the seventh year, thus he is violating two explicit prohibitions. We can have similar situations with obligations. We may end up in a situation where we have conflicting obligations.So we may have, for example

1. An obligation to do A
2. A prohibition to do A
3. Another obligation (for a different reason) to do A
4. Another obligation not to do A.

How do we represent and handle such a situation?

According to our model such situations exist only in the normative plane, in the language of the operators O_T and F_T. To indicate possible different source we need indeed 248 different O_Ts and 365 different F_Ts, to reflect what exists in the Bible.

So we represent the above situation as

1. $O_{(1,T)}(A)$
2. $F_{(1,T)}(A)$
3. $O_{(2,T)}(A)$
4. $F_{(3,T)}(\neg A)$

For the practical level, what one is actually supposed to do in any given situation? We make a decision and represent this by one operator O_D. If the decision in the case of (1)–(4) above is to do A, we write $O_D(A)$, and if the decision is not to do A, then we write $O_D(\neg A)$.

The decision what to do is done by Tables 5 and 6 and Remark 11.

2.8 Comparison with legislation in law

We have already remarked that the general characteristics of Talmudic legal system is different from the general ones. This follows mainly from the fact that in general legal systems we do not have obligations in the sense of previous subsections.

We may have laws about what to do and what not to do but not in the normative sense of previous subsections. An ordinary legal system does not give reward for acting according to the obligations of the law; it only imposes punishment on violations of the law. Therefore there cannot be any serious distinction in general law between violation of an obligation to do something and a violation of a prohibition to do something.

In contrast, in the Talmud, such distinctions are central. An obligation is a command to better yourself and a prohibition is a command to stop yourself from deterioration, as discussed in Section 5.1. It is therefore natural that there would be different status to obligations and to prohibitions. God rewards you for obeying his obligations and either God or a local court will punish you for violating a prohibition. This is also why the Talmud requires special rules in cases of conflicts between obligations and prohibitions. We do not get too many such rules in general legal systems.

The distinctions between the normative level (O_T, F_T) and the practical level (O_D) is of course applicable also in general ethical and legal systems, and may even help resolve some legal paradoxes.

For more details, see our book [6].

2.9 Comparison with preference based models

This section will compare our results with three central papers on preference based models for obligations, namely Carmo and Jones [39], Cholvy and Garion [38], and Tore and Tan [106].

To do this successfully and avoid a lengthy presentation of the theories of these papers, we chose an example addressed by all and use it to show the differences between our section and their papers.

Consider the dog example:

(a.) There ought to be no dog.

(b.) If there is no dog, there ought not be a warning sign.

(c.) If there is a dog, there ought to be a warning sign.

We begin by listing the models we can form out of the propositions "dog" and "sign". These are:

$$\begin{aligned} w_1 &= \text{dog} \wedge \text{sign} \\ w_2 &= \text{dog} \wedge \neg\, \text{sign} \\ w_3 &= \neg\, \text{dog} \wedge \text{sign} \\ w_4 &= \neg\, \text{dog} \wedge \neg\, \text{sign}\ . \end{aligned}$$

A preferential model for obligation will give a preference relation on the worlds and will derive the obligations from the preference.

Let $w \gg w'$ mean w is better than w'. For example, something reasonable compatible with (a)–(c) is:

$$w \gg w_1 \gg w_2$$

and

$$w_4 \gg w_3.$$

We can argue about what what the status of w_3 is in relation to w_2 or w_1.

The Talmudic view, as we saw, is looking at what world state we are, say w_2, and checking what obligations and prohibitions are activated in this state. In this case it would be no dog and yes sign. The Talmud gives rules to decide what to do in practice. The obligations and prohibitions activated at w_2 may be conflicting. (We might remove the dog and put up a sign!)

The preferential approach would largely ask the agent to move to a better preferred world if he can. Carmo and Jones, for example, distinguish between ideal obligation O_i (our O_T?) and actual obligation O_a (our O_D?).

There is another feature put forward by Carmo and Jones and others and this is the question of whether the agent controls the possibility of change? The agent may not be able or willing to remove the dog or put up a sign. The Talmud recognises this possibility but does not take it into consideration in connection with the question of whether there is a violation. The Talmud always counts as violation if the state is not as it should. So if there is a dog and no sign there are violations. The Talmud might say if the sign is too expensive then the obligation is not valid, but if signs are not available at all then there is still violation.

Carmo and Jones and, to some extent, Cholvy and Garion, may say that if the agent is unable to execute an obligation then either there is no obligation any more or maybe at least there is no violation. The Talmud does not make such connections.

The Torre and Tan paper presents a system of obligations based on preference and defines

$$\alpha \text{ should be done if } \beta \text{ is done is true iff}$$

1. No $\neg\alpha \wedge \beta$ state is preferable to $\alpha \wedge \beta$ state and
2. The preferred β states are α states.

The Talmud does not use preferences to define its obligations but decrees 613 types of prohibitions and obligations.

To sum up: The above discussion shows that the preferential approach is completely different in flavour from the Talmudic approach. The preferential approach wants the agent to move to a better preferred world. So we need to look how the worlds are organised, see where we are and decide where to go.

The Talmudic approach gives obligations and prohibitions which are triggered by the state of the world you are in. These may be conflicting. There are rules to tell you what to do. You are not moving to a better world but making yourself better.

3 CONTRARY TO TIME CONDITIONALS IN TALMUDIC LOGIC

3.1 Preliminaries

We consider conditionals of the form $A \Rightarrow B$ where A depends on the future and B on the present and past. We examine models for such conditional arising in Talmudic legal cases. We call such conditionals Contrary to Time (CTT) conditionals.

We also consider, in a continuation paper, lack of clarity in actions arising from assertions of the form $B((ix)A(x))$ where A depends on the future and B on the present.

Three main aspects will be investigated:

1. Inverse causality from future to past, where a future condition can influence a legal event in the past (this is a man made causality).

2. The status of identification of entities in the present using definite descriptions involving the future.

3. Processes which create reality via legal decisions and norms.

We shall see that we need a new temporal logic,which we call Talmudic Temporal Logic (**TTL**) with linear open advancing future and parallel changing past, based on two parameters for time.

3.2 Introduction and motivation

We motivate our system through some typical example and then we present a two dimensional temporal logic to represent them. We continue with some real examples from the Talmud.

EXAMPLE 14 (Sale of a company). Consider the following scenario for the sale of a company: Smith sells his business to Jones on 1 January 2010. The exact amount for the sale (say x) will be determined by the company's performance during 2010. So x is defined to have value on 1 January 2010 according to the profits and growth as announced and calculated on 31 December 2010 for the period 1 January–31 December 2010. Say if growth and profits exceed 10%, the price is $x = k$ pounds, otherwise $x = 0.8k$ pounds. Such conditions are common in business.

We have to be careful here to distinguish whether the agreement is that the purchase price is $0.8k$ and an additional $0.2k$ is paid if performance exceeds 10% or whether the performance determines the purchase price retrospectively in time. The difference may matter for the purpose of taxation.

Smith and Jones agree that if growth is less than 1% then the sale is off, i.e. the question of whether on 1 January 2010 the company is sold or not depends on future performance. If the performance is low, then there is no sale.

REMARK 15 (Discussion of Example 14). We need to address the following in view of Example 14

1. Adequate logical language to describe such phenomena, and all relevant distinctions. We call this area Contrary to Time (CTT) conditionals.

2. Identify exactly our intuitions and options for evaluating such cases.

3. See what extensions to the language in (1) are needed to model legal rulings for such cases. (This item will become clearer later in Section 3.)

There are definitely legal points which need to be clarified for this case; for example, we may ask who owns and runs the company during the year 2010? Suppose we agree that Smith continues to run the company. Suppose Mr Smith is negligent in running the company during 2010 and as a result of his mismanagement the company incurs serious losses. Can Mr Jones ask for compensation? The loss of a good company he bought? Smith can claim that there was no sale! However, had the company been sold outright on 1 January 2010 and Mr Smith was just acting as interim director, then certainly he would have been found negligent and been asked to pay compensation!

Let us compare the above sale example with a different example.

EXAMPLE 16 (The Princess's Marriage). The King of a certain kingdom has a beautiful daughter ready to be married. To find her a husband, her father made a call to all young princes from other kingdoms to come and compete for her hand in marriage. Many came. The King said that the bravest prince who wins the test shall marry her. The test is to overcome and kill the big monster known to reside in the mountains. Some princes said that perhaps they are not qualified in bravery for such a competition. They asked "can other members of their family do the killing for them"? The King agreed that if a brother or father of the candidate does the job and slays the monster then the candidate wins.

The King was very careful in setting up this competition. He made each candidate give his daughter a ring and sign a copy of a proper set of marriage papers with the condition that the marriage becomes immediately valid if the candidate slays the monster in the future.

The candidates went to the mountains looking for the monster. Many perished in the mountains, and many were killed by the monster. Eventually one of them came back after many months with the head of the slain monster. This candidate (actually it was the brother of the candidate which qualified the candidate) now claimed the validity of the marriage and declared himself married to the princess from day 1! He was told that during all these months the princess fell ill and died!

The prince said that he is now a widower, since the marriage was valid retrospectively, and that he inherits all the princess's estates.

There are questions to be clarified about this story.

1. What is the status of the princess before the monster is slain? I.e. from the time she signed the marriage papers until the time someone killed the monster?
2. What is the status of the princess in the period if no one slays the monster? (All candidates perished.)
3. What is the status of the princess in the period if two princes cooperated and slay the monster? (The princes may have been in a position where they had to cooperate to survive an attack by the monster.) Who is married to the princess now? None?
4. If the princess died in the period, should not the deal be immediately cancelled?

Clearly we need a formal language to talk about such examples and represent what is involved.

Let us tentatively write

$\alpha(x) = x$ is a prince qualified to compete for the princess.

$\beta(x, y) = y$ is a brother or father of x

$\varphi(t, x, m) = x$ kills the monster at time t.

$\Psi(s, x, y) = x$ marries y at time s.

EXAMPLE 17 (Insurance example). We now give a common day to day example of insurance. A car policy can be viewed as a temporal statement of the form:

$\mathbb{P}$: car stolen $\Rightarrow$ we pay

where by 'stolen' we mean any other damage as well.

The policy has validity of 12 months, say 1.1.2010–31.12.2010. Let us represent this by

$$[1.1.2010\text{–}21.12.2010] : \mathbb{P}$$

During January 2011, the policy needs to be renewed. The usual practice is that the policy can be renewed any time in January 2011 and the validity of the renewal policy would be from 1.1.2011–31.12.2011. So assume that the date of renewal (i.e. payment for the year 2011) is 20 January 2011. What is the situation of insurance coverage on 15 January 2011?

Assume the customer had no intention of renewing his policy. So he was not going to pay anything during January 2011. Then on 15 January 2011, the car was stolen. He needs to hurry and pay the premium by 31 January 2011 and his car will be insured on 15 January 2011. The insurance company cannot say that he paid because the car was stolen and otherwise he would not have paid. This is irrelevant.

So we see again we have three periods here, as in Figure 1.

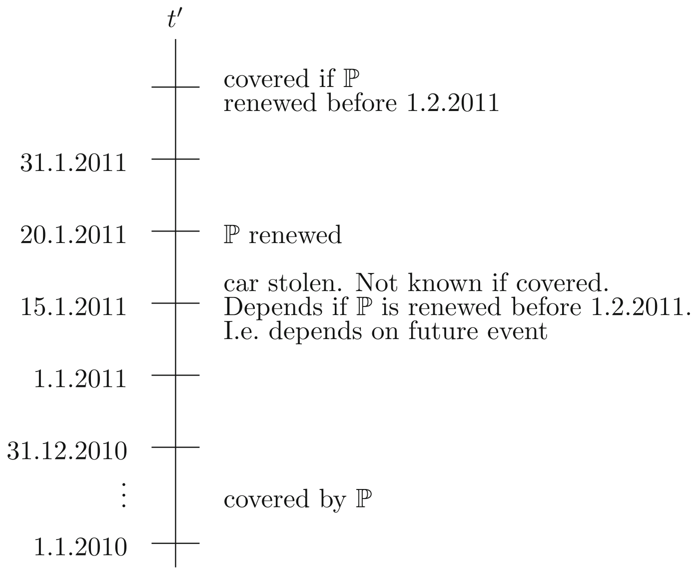


Figure 1.

3.3 Language for contrary to time with examples

3.4 Choice of language

There are in the literature two main options here:

1. Semantically based languages where we have a flow of time with branching future and a temporal language to talk about it. This we call **Traditional Temporal Logic**.

There are many variations of such systems, depending on the flow of time, the operators used and the number of time indices used.

2. A syntactically based language where the future does not exist yet and we only have various syntactical formulas talking about it. The future formulas can be made true by our actions but there is no future (branching or not) in which they can be semantically evaluated.

   A future statement of temporal logic can be understood in two ways: the declarative way, that of describing the future as a temporal extension; and the imperative way, that of making sure that the future will happen the way we want it. Since the future has not yet happened, we have a language which can be both declarative and imperative. This we call **Executable Temporal Logic**.

Our own new TTL will emerge as a new third approach, capable of handling backwards causality in time. It is a completely new approach, but can be viewed as some sort of a variation of the executable family.

We now give more details on (1) and (2) above.

Traditional Temporal Logic

When deciding on how to model temporal phenomena, the first step is to decide on the flow of time. In most applications time is taken to be acyclic (non circular) with the past linear (no ambiguity or branching of the past) but with the future being branching to allow for the fact that the future is not determined.

We start with a flow of time $(T, <), T$ is the set of moments of time and $<$ is the earlier later relation. We have that $(T, <)$ satisfies the following axioms.

1. $<$ is transitive and irreflexive.

2. For every t, the set $T_{<t}$ is linearly ordered by $<$

$$T_{<t} = \{x | x \leq t\}.$$

Figure 2 shows that $(T, <)$ can be a tree, branching into the future with linear past.

Additional axioms on $(T, <)$ are needed to ensure the tree property. For our purpose, we need not insist on trees.

There are two points of view we can adopt when considering the flow of time.

1. The external view, where we stand outside time and look at the entire history like God viewing the history beneath us. In this case there is no fundamental difference between future and past. There is only one

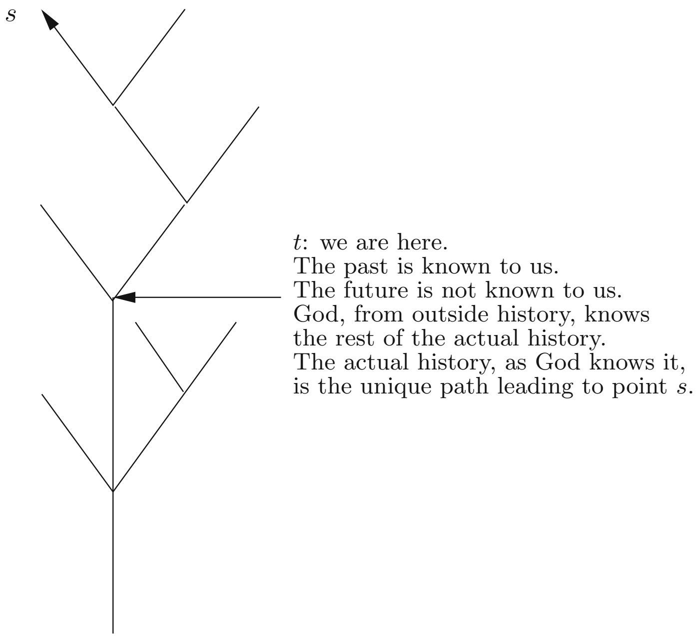


Figure 2.

actual real linear flow of time , the real history as it happened, and if we have a branching flow then the real history must be marked in the flow as what actually happened. See Figure 2.

2. The internal view, where we see ourselves as ordinary mortals residing at some point in history, and the future is truly branching, because it has not happened yet, and all we have is our linear past.

Once we decide on how we view history , we need a language to talk about it. Again we have two options:

OPTION 1: Use a global language to talk about time in absolute sense. This is like using a global clock dates and saying for example

> In the year 701 BC , Sennacherib king of Assyria attacked all the fortified cities of Judah and captured them

OPTION 2: We use temporal markers relative to where we are (e.g. tomorrow, yesterday, when we got married) or markers of any kind. This is how the Bible does it:

> Kings 18:13-15 In the fourteenth year of King Hezekiahs reign, Sennacherib king of Assyria attacked all the for-

> tified cities of Judah and captured them. So Hezekiah king of Judah sent this message to the king of Assyria at Lachish: I have done wrong. Withdraw from me, and I will pay whatever you demand of me. The king of Assyria exacted from Hezekiah king of Judah three hundred talents of silver and thirty talents of gold. So Hezekiah gave him all the silver that was found in the temple of the LORD and in the treasuries of the royal palace.

OPTION 3: Use a mixed approach, as may be convenient for the application at hand.

We shall use Option 3.

We consider a two sorted predicate logic with atoms of the form

$$P(t, x_1, \dots, x_n).$$

The first coordinate variable t is a time variable ranging over $(T, <)$. The other variables $x_1, \dots, x_n$, range over a variable domain D_t. We consider such $P(t, x_1, \dots, x_n)$ as facts. We can quantify over t and x_i, and write formulas like

$$\forall x \varphi(t, x) \text{ for } \forall t \varphi(t, x).$$

It may be the case that some atomic predicates are independent of time. In this case we write $\alpha(x_1, \dots, x_n)$ without a time variable and understand that they are either true for all time or are false for all time. So if $\forall t P(t, x) \vee \forall t \neg P(t, x)$ holds we can write it as $P(x)$.

We allow for the *Iota* operator $(ix)\varphi(x)$ and the $(it)\varphi(t)$, meaning

"the unique x (or t) such that $\varphi(x)$ (or $\varphi(t)$) holds."

Iota can be used only when such a unique element can be proved or assumed to exist.[53]

Let the above language be called temporal $\mathbf{L}$. It is a two sorted language. The atomic wffs of $\mathbf{L}$ are either $P(t, x_1, \dots, x_n)$ or $t < s$.

[53]We can formulate a formal language for the Iota based on the temporal logic of the current section enriched with the Iota symbol. We shall have to allow for $(ix)\phi(t, x)$ and for $(it)\phi(t, x)$ to be formed independently of any semantical condition. This means that we have to give denotation to these expressions also in the case where there exists more than one element satisfying ϕ and for the case where there exists no element satisfying ϕ.

There are several options in the literature of what to assign to the Iota expression in such cases, but as far as we know, there is no discussion in the context of temporal logic. The classical options are to make the Iota expression undefined or to assign an arbitrary element to it. In temporal logic it is better to view the Iota elements as a non existent free element which may come into existence, should a unique element show up.

There is no need for us to pursue this course of action. It is too complex. The continuation, Section 3.8 avoids the use of the Iota by using quantum superposition models.

L is closed under the classical connectives and the quantifiers $\forall t$ and $\forall x$, and the *Iota* operator $(ix)\varphi(x)$.

A model for **L** is a flow of time $(T, <)$ and a classical domain D_t associated with each $t \in T$.

We have an assignment h giving for each atom P a set of tuples $(t, x_1, \ldots, x_n)$ where $x_i \in D_t, i = 1, \ldots, n$.

For simplicity, let us assume that $D_t = D$ for all t. This assumption simplifies the model because we do not have to deal with $\varphi(t, x)$, when $x \notin D_t$. Of course, in real life, people are born and die and the domain changes. But also in real life we talk about dead people and yet unborn children and we refer to them and interact with "them" and do things with them and so we can allow for a predicate $\lambda(t, x)$, saying x is alive in time t. So x can be either live or a dead person or a person to be born.

Our basic statements have the form

$$s : \varphi(t_1, \ldots, t_k, x_i, \ldots, x_k).$$

reading

at time s $\quad \varphi(t_1, \ldots, t_k, x_1, \ldots, x_k)$ is claimed to hold.

φ a complex formula, e.g.

$$\varphi = P_1(t_1, x_1) \wedge P_2(t_2, x_2).$$

This gives rise to a two dimensional logic.

Note that we use two indices for time. The t_1, t_2 time indices are according to Option 1 and the s index which is the second dimension, is according to Option 2.

Time s, the second dimension, is where we are (Option 2) and from time s, we are talking like gods (Option 1) about other times t_i.

First note that we have to assume that for atomic sentences, all observers at s agree on the past. Thus to express this formally we need to make the assignment h depend on s. Thus $h_s(P)$ gives a set of tuples $(t, x_1, \ldots, x_n)$ meaning

According to s, $P(t, x_1, \ldots, x_n)$ holds.

Our condition becomes

- $(t, x_1, \ldots, x_n) \in h_s(P)$ and $s \leq s' \Rightarrow (t, x_1, \ldots, x_n) \in h_{s'}(P)$.

h_s may not agree on future predictions.

So if $s < s' < t$ we may have

$(t, x_1, \ldots, x_n) \in h_s(P)$ but
$(t, x_1, \ldots, x_n) \notin h_{s'}(P)$.

Executable temporal logic

This is best explained with an example.

EXAMPLE 18 (Simplified Payroll). Mrs. Smith is running a babysitter service. She has a list of reliable teenagers who can take on a babysitting job. A customer interested in a babysitter would call Mrs. Smith and give the date on which the babysitter is needed. Mrs. Smith calls a teenager employee of hers and arranges for the job. She may need to call several of her teenagers until she finds one who accepts. The customer pays Mrs. Smith and Mrs. Smith pays the teenager. The rate is £10 per night unless the job requires overtime (after midnight) in which case it jumps to £15. Mrs. Smith uses a program to handle her business. The predicates involved are the following:

$A(x)$ x is asked to babysit
$B(x)$ x does a babysitting job
$M(x)$ x works after midnight
$P(x, y)$ x is paid y pounds.

In this setup, $B(x)$ and $M(x)$ are controlled mainly by the environment and $A(x)$ and $P(x, y)$ are controlled by the program. We get a temporal model by recording the history of what happens with the above predicates. Mrs. Smith laid out the following (partial) specification:

1. Babysitters are not allowed to take jobs three nights in a row, or two nights in a row if the first night involved overtime.

2. Priority in calling is given to babysitters who were not called before as many times as others.

3. Payment should be made the next day after a job is done.

Figure 3 is an example of a partial model of what has happened to a babysitter called Janet (note that time is discrete and flows upwards). We have a record of the past activity. The future is open ended but we read the future specifiation as instructions to us to execute the appropriate actions which make it true. If we always succeed then we get a temporal flow which satisfies the specification in the traditional sense.

In any case we get a traditional model. This model may or may not satisfy the specification. In general we would like to be able to write down the specification in an intuitive temporal language (or even English) and have it automatically transformed into an executable program, telling us what to do day by day.

3.5 Examples

We now give some examples of the use of time in the Talmud.

7 $\sim A(J), \sim (J), \sim M(J)$
6 $A(J), B(J), M(J)$
5 $A(J), B(J), M(J)$
4 $\sim A(J), B(J), \sim M(J)$
3 $A(J), B(J), M(J)$
2 $A(J), B(J), M(J)$
1 $A(J), B(J), \sim M(J)$

Figure 3. A model for Janet

EXAMPLE 19. Consider:

> s: x enters the room at t.

written

$$s : E(t, x, \text{ room}).$$

Reading: At time s, it is said that the element x entered the room at time t. we distinguish two cases:

Case 1 $t \leqq s$. This is a statement about the present or the past. It has no prediction.

Case 2 $t > s$. This is a statement about the future.

We may use, at time s, the statement $E(t, x, \text{ room})$ to identify a person at time s and say something about this person.

We can say

> s: the person x who (will enter) entered the room at time t, is now (at time s) in prison.

Using the *Iota* symbol, we write

$$s : P(s, (ix)E(t, x, \text{ room}))$$

To be able to say that and use the *Iota* symbol we need the condition that exactly one person entered the room at time t.

If $t > s$, we are identifying the person by what is going to happen to him. So at time $s, s < t$ we cannot yet be sure whom we are talking about. We can only be sure at time $s < t$ that exactly one person will enter the room at time t. We are saying at time s that whoever this person is, he is the one we are talking about.

Some people may take the view that identifying x by a future event at time t is not acceptable for some purposes (e.g. legal inheritance documents, etc). The Talmudic approach to such examples will be addressed in detail in [93].

EXAMPLE 20. Consider the following contract $\Psi(s,a,y)$, between individual a and individual y. Assume this document is put forward for legal approval at time s. The contract has the form $s:\Psi(s,a,y)$. It is a contract at time s between a and y. The content of the contract is spelled out by Ψ. y is a variable for an individual to be identified as follows:

$$y \in Y = \{y|\alpha(s,a,y)\}.$$

$\alpha(s,a,y)$ is a predicate identifying a set of ys. If α is timeless and does not depend on s, then we write $\alpha(a,y)$.

Furthermore from among this set of ys we further identify those ys which stand in relation β to a specific z_0. If this relationship depends on s we write $\beta(s,z_0,y)$ If this relationship is timeless we write $\beta(z_0,y)$. This z_0 is identified by future time t, i.e. $z_0 = (iz)\varphi(t,a,z)$. Let us assume that both α and β are timeless. So we have altogether

$$\alpha(a,y) \wedge \beta((iz)\varphi(t,a,z_0,y).$$

It may be that α and β and φ depend on other elements $b_1,\ldots,b_k$, with possibly $b_1 = a$. In such a case we write $\alpha(s,b_i,y)$ and we write $\beta(s,b_i,z,y)$, and also $\varphi(t,b_i,y)$.

The Talmud uses $\alpha,\beta,\varphi,\psi$ in two main forms:

1. **Conditional form**

   $s:\varphi(t,b_i,y_0) \Rightarrow \Psi(s,b_i,y_0)$
   and
   $s:\neg\varphi(t,b_i,y_0) \Rightarrow \neg\Psi(s,b_i,y_0)$

   where $t > s$ and where y_0 is a certain named individual and where $\Rightarrow$ is a possibly nonclassical strict or resource implication or an intuitionistic constructive implication. We require $\Rightarrow$ to satisfy modus ponens

   $$A, A \Rightarrow B \vdash B$$

   In Hebrew the name for this is “Tenai”, meaning “condition”.

2. **Choice form**
   (In Hebrew the name is “Breira”, meaning “choice”.)

   The form is

   $$\Psi(s,b_i,y) \equiv \alpha(b_i,y) \wedge \beta(b_i,y,(iz)\varphi(t,b_i,z))$$

where y is a variable, $t > s$ and the predicate $\varphi(t, b_i, z)$ defines a unique z at time t, unique because $\exists! z\varphi$ is assumed to hold, or hoped that it will hold!

So what is happening here is that $a = b_1$ says I want to enter in contraction relation with an element $y \in Y$.

The identity of this element is determined by φ. β says what is the nature of the relationship that a enters with each possible y.

Both β and φ may depend on additional elements b_i which may include $b_1 = a$ or not, i.e. we have $\varphi(t, b_i, z)$ and $\beta(z, b_i, y)$. For the purposes of comparison, note that if the conditional is expressed in classical logic we get
Classical logic conditional:

$$\Psi(s, b_i, y_0) \equiv \varphi(t, b_i, y_0).$$

Parameters of importance here are the following:

(a) does Y allow for more than one element?

(b) Is $t > s$ or $t \leq s$? How is t defined? It may be defined or specified as $s : \gamma(t, x_i)$ in which case is there a unique t_0 known at s or only known at some later t, $s < t < t_0$?

(c) Is a one of the b_i?

Here is an example:

EXAMPLE 21. a says "I will sell to one of my cousins either my Montblanc pen or my Parker pen".

The pen I sell depends on the cousin and on a third party b. There are two cousins John and Mary. If I sell to cousin John, it will be the Parker pen. If I sell to cousin Mary then which pen I sell depends on party b. The dependence on b is whether b wins the election tomorrow. The cousin of choice for sale is the one who calls me first tomorrow to ask about the sale. Here $\alpha(a, h)$ is

"y is cousin to a" (i.e. $y =$ John or $y =$ Mary).

$\beta(z, y)$ is

"a sells Montblanc to z if b wins and $z =$ Mary and a sells Parker to z if z is cousin John".

Further detailed discussion of problems arising from actions involving elements identified by future properties of future actions is done below.

3.6 The legal decision functional

In Section 2 we used temporal logic built up from classical logic with variables and constants over time. We use $<$ for the earlier–later relation and form atomic predicates of the form

$$P(t, x_1, \ldots, x_n)$$

where t is time and x_i are individuals in the domain. We consider $P(t, x_1, \ldots, x_n)$ as facts. We ask that $<$ is linear acyclic in the past and branching in the future.

So $<$ is transitive and irreflexive, and anti-symmetric.

This makes it tree like.

The basic logical unit is

$$s : A(t_i, x_j)$$

reading:

> at time s the statement $A(t_i, x_j)$ is considered to hold true.

This reflects the view that we reside within history and not outside it, and at each moment of time s, we have our views about both the future and the present and past!

So to give examples:

> John gave Mary the book at time t stated at time s is translated as
>
> $$s : G(t, J, M, \text{ book})$$

The linearity of the past implies that there is only one version of history at any time r and they all agree.

$$r : G(t, J, M, \text{ book})$$

for $r > t$ implies that for any $r' \geq r$, also $r' : G(t, J, M, \text{ book})$.

Let $(ix)\varphi(x)$ be definite description operator

> "The x such that uniquely $\varphi(x)$ holds of x".

Then we can say at time s:

> John bought at time s the book he gave to Mary at time t.

$$s : \text{ Buy}(s, J, (ix)G(t, J, M, x))$$

The above can be problematic if we have $s < r < t$.

Today we say:

> John bought yesterday the book he is going to give to Mary tomorrow.

If John took three books yesterday from the bookshop, paid for one only, promised to look at the books, make a decision and return two tomorrow after he gives one to Mary, then we are not able to identify which book he actually bought yesterday until tomorrow.

So we have

$$r: \ \mathrm{Buy}(s, J(ix)(G(t, J, M, x)).$$

If $r < s$ or $r < t$ we have a problem.

This problem is legal.

1. One might adopt the view that there is no sale at time s until John gives his chosen book to Mary at time $t, s < t$.

2. One might adopt the view that there is a sale of a book at time s, but the identity of the book is not known until time $t > s$.

3. One might take a "quantum view" and say that there is a sale of all three books at time s (like a superposition in quantum mechanics) and that two of them are returned at time $t > s$.

4. One may take the view that you cannot make a sale in this way and that a specific book needs to be put on record as sold at time s and that if John wants to give a different book to Mary at time $t > s$, then the record will show that he exchanged the book that he bought at time s for another book.

The differences between these views may have consequences for taxation. Say, for example, the payment of VAT.

If t is much later than s, then the question of how many books were sold at s becomes relevant to the bookshops VAT report.

The seller may adopt option 1 — no sale, while the VAT people may adopt option 3 and claim VAT for three books.[54]

Let us be more formal now: At time r, the statements $A(t_i, x_j)$ are descriptive. They say what happens to x_j, t_i over time. $A(t_i, x_j)$ can be written as a Boolean combination of pure sentences of the form

$$A_k^{>r}, A_k^{=r}, A_k^{<r}, k = 1, 2, \ldots$$

where $A_k^{>r}$ talks *only* about the future time (of r) and $A^{<r}$ talks only about the past of r and $A^{=r}$ talks only about time r.

We also have normative sentences, like

[54]In the UK you pay VAT the minute you issue an invoice. The way businesses go around this rule they send a "reminder" and write on it in bold letters "This is not a VAT invoice". I am always tempted not to pay, claiming I have never been sent an invoice!

> Does Mr Jones own the company on 2 January 2010, having bought it on 1 January 2010, but with a condition that the sale is off if the company incurs losses during 2010?

We need a meta predicate (or a functional) which normatively declares what Mr Jones can and cannot do.

The normative sentences are needed because much of our day-to-day reality is man (legally and normatively) made. It is not just facts but agreed statutes.

Let $\mathbb{H}$ be such a predicate. Let $\mathbf{e}_1, \ldots, \mathbf{e}_m$ be possible normative states which various formulas can be said to be in. Thus at each time s, we can look at statements of the form $s : B(t_i, x_j)$, where t_i are maybe past or future to s, and where B needs a normative decision. We can write:

$$s' : \bigwedge_k A_k(t_i^k, x_j^k) \to \mathbb{H}^{s'}(s : B(t_i, x_j)) = \mathbf{e}_m.$$

Note that the predicate $\mathbb{H}$ is in an Option 1 language talking about both dimensions. $\mathbb{H}$ is indexed by a third dimension.

Formal definitions will need to be given in a later section.

EXAMPLE 22. Time is today.

A = yesterday Smith bought a company from Jones under the condition that if tomorrow the shares are down from today, the sale is off.

B = we ask: can Jones sign contracts for the company without the consent of Smith?

$\mathbb{H}$ might say **yes, no, indeterminate**. So we write

$$\text{today} : A \to \mathbb{H}^{\text{today}}(B) = \ \mathbf{no}.$$

3.7 Talmudic examples

We now give some examples from the Talmud.

EXAMPLE 23 (Sale of property). A man a sells his property to a buyer b on the condition that when he (the seller) has more money, he can buy the property back from b. The ruling is that the sale and the condition are valid and if b refuses to sell the property back to a then the sale is annulled from the start. (Origin of ruling: *Shulchan Aruch, Choshen Mishpat*, section 182:512.)

EXAMPLE 24 (Divorce variations). The following examples illustrate possible logical connections between the original action and the condition it relies on

1. **Physically possible but legally forbidden condition**
   A man a signs divorce paper to his wife b on the conditions that she has sex with another man c.

The ruling is that if she does have sex with c then the divorce is valid.

Note that there is a fine point here. It is sinful to have sex with c while she is married to a and therefore she cannot fulfil the condition without initially sinning. Once she has sex with c then she is retrospectively divorced and therefore there is no sin.

Compare with Example 26 below. (Origin and ruling in *Jerusalem Talmud and Shulhan Aruch, Even Ha-ezer*, 153-518.)

2. **Legally impossible condition**
   In *Gittin* 84–1, there is a variation of the above example, where a gives divorce to his wife on the condition that she marries c (rather than the condition of having sex with c). In this case the ruling is that she can get married to c and the marriage is legal.[55]

   The puzzling question arises that when she wants to marry c, the condition is not fulfilled yet and so she is still married to her husband a and so she cannot have a legally valid marriage to c. So how can she ever fulfil the condition? And why the ruling is that she can marry c?

   The answer is that the condition itself, because of the above considerations, is not legally consistent and so she was originally divorced without (the legally inconsistent) condition.

   Thus if an action is taken with an inconsistent condition it is deemed that the action was taken without that condition , as opposed to ruling that no action is taken (because there is something wrong with the condition).

3. **Logically looping condition**
   In *Gittin* 83–1 we have the example that a divorces his wife on the condition that she does not marry c.[56]

   As long as the woman does not marry c, her divorce from a is valid and she is divorced and can marry anyone she wants. If she marries c however, she violates the condition and therefore her divorce from a is not valid retrospectively, and so she is still married to a. But then, if this is the case, her marriage to c is not valid because she is still married to a!

   But if her marriage to c is not valid then she has not violated the condition of her divorce from a, and therefore she is indeed divorced and therefore her marriage to c is valid, etc., etc., and we are in a loop!

[55] In practice such condition is not allowed because it looks like legalised wife swapping.
[56] c may have been b's "friend", and the husband insisted on this condition.

4. **Condition that one of the partners dies**
   These are cases where a gives divorce to b on the condition that x dies, where x is one of $\{a, b\}$.

   Such circumstances arise when for example a is a soldier who might die in battle and would like to ensure that in the event of his death, his wife b is retrospectively free from the time he has left for the battle. This was the custom in Biblical times, as reported in *Shabat* 56-1 The case where the condition is the death of the wife is in *Yoma* 13-1.

   In both cases the ruling is that the divorce is valid

   There is also a lengthy detailed discussion of such cases in *Gittin* 72-1 and in *Kidushin* 60-1.

   In this discussion there is a need to clarify whether some specific "conditions" fall under the case of backwards causality conditional actions or cases of the use of the Iota operator.

In Biblical times it was possible to marry more than one wife. To divorce a wife one has to write divorce papers specifically intended for the wife to be divorced. One cannot fill out ready made divorce forms. So the example 25 below contains a doubt arising from a future condition which is used to identify the wife.

EXAMPLE 25. At time s, a writes divorce papers intended for his wife Rachel. He has two wives, both named Rachel, and he says he will decide which one to give it to tomorrow. He makes his decision dependent on one of the following:

1. stock market goes strictly up or not, or
2. whether he plays chess tomorrow, or
3. whether Rachel wears brown shoes tomorrow.

Let $A = a$ writes at time s divorce papers for wife $(ix)(B_k(s+1, x, a))$ where $k = 1, 2, 3$ according to the three cases above.

Question. Are the divorce papers valid?

Analysis. B_1 is a clear cut definition identifying which one of the two Rachels is meant in the document, but the defining predicate depends on the future. The future is not dependent on him.

B_2 is like B_1 except that the future is also dependent on him. He can play or not play chess under normal circumstances.

B_3 may not even be a legitimate definition of both wives wearing brown shoes or neither.

Furthermore, today, as far as we know, any one of the two Rachels may end up divorced!

There is some degree of indeterminacy.

We can have a policy in such cases. The following are some options:

1. To be valid at r $(ix)B(t, x, a)$ must depend only on the present or past of r, i.e. $(t \leq r)$.

2. We do not mind about B as long as it is objective (a not in B).

3. We don't mind at all. Any B will do.

4. Any of the above cases (1)–(3) makes the answer to the question doubtful. So we do not know (we are in doubt) whether the divorce paper is valid. However, we can be strict and say that we deem it that there is no divorce in any other context where the question of Rachel's divorce is involved.

   So Rachel is not considered divorced because of the doubt but she is not available to a Cohen[57] either, because we deem Rachel as divorced for that context, again because of the doubt.

3.8 Introducing temporal Talmudic logic (TTL)

In this section we discuss our advancing future–changing past model, capable of handling conditionals.

We begin with methodological remarks, on how we discovered Talmudic Temporal Logic. There are many examples and discussions in the Talmud about various cases of conditional actions and various legal rulings about them. This is our body of evidence. We were looking for a temporal model which can accommodate and explain all the different approaches and views of the Talmudic scholars discussing and ruling about all of these examples. What we call **TTL** is the simplest such model which can do the job.

What do we mean by a logical model? One's immediate reaction might be just to give syntax and semantics for the appropriate language and define logical consequence using the semantics. However, this is not sufficient for two reasons:

1. The Talmudic data by nature is a body of arguments, counter arguments and debates and so we would expect modelling using a proof theoretical and arugmentation system which can also model the way the debates are executed.

[57]A Cohen cannot marry a divorced woman, only a widow. So if Rachel's husband dies, is she deemed for the purpose of marrying a Cohen as already divorced, or just a widow?

2. The nature of the temporal examples, as we shall see below, involves alternative histories which then disappear as time goes on, and so even if we give a many dimensional parallel histories semantic model, the natural way of defining a semantic consequence will be too weak to reflect what is really happening unless we are able to add to the language syntactical constructs explicitly talking about alternative histories. Such constructs, however, are not present in the Talmud.

In this section we introduce **TTL** using modern examples. The next section will present **TTL** as a formal logic and later we show how this model explains Talmudic examples. A very detailed discussion can be found in our companion book [8].

We now need to clarify some concepts. Consider the following statements.

1. The vase is broken into two pieces.
2. Mary is married.
3. John's income is from employment on a sea-faring ship. (Therefore is tax free!)

Statement 1 is a physical statement. It is not a legal or social convention statement. One has a bit of leeway in understanding what "two pieces"means and if one piece is very small we might say that the vase is "chipped", not broken. We might even argue, in the case of a slightly bigger piece, whether we can still say "chipped", or say "broken". The difference may be important for insurance purposes. Do we replace the vase or do we fix it? It may even be the case that the insurers stipulate that "broken" means, as far as they are concerned, broken into 3 pieces or more, but now we are into the legal domain.

Statement 2 is a statement of legal and social agreement. Society and the law allows for a marriage action **a** to take place, provided certain preconditions $\mathbb{C}_{\mathbf{a}}$ do hold and the result of which we get the truth of the legal predicate x is *married to* y. x *is married* is $\exists y(x$ *is married* to $y)$.

Being a legal predicate, the marriage status can be changed by a divorce action **b**. Again, given preconditions $\mathbb{C}_{\mathbf{b}}$ we can make true the predicate x *Divorce* y. We have that if x *Divorce* y holds at time t, then $\neg(x$ *is married to* $y)$ holds at t.

This is different from "broken" predicate. No action can "$\neg$ break". We can "glue" the pieces but what we get is "broken but glued" or "broken (leg) but healed", etc.

Statement 3 is also a legal statement. Here one can legally change the meaning of "employment on a ship". In fact the British definition included working on an oil rig. To increase taxation the government changed the meaning into "moving ship", thus excluding oil rigs.

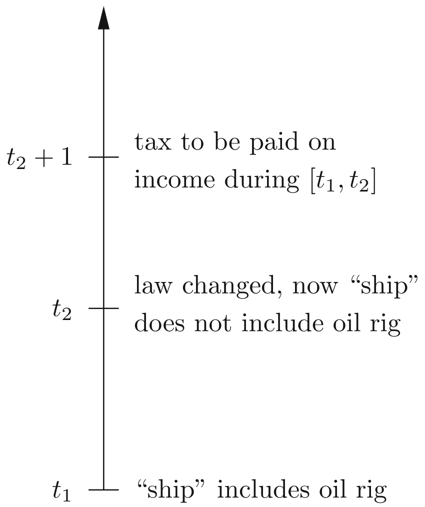


Figure 4.

Such a move can cause a lot of resentment because it involves backward taxation as Figure 4 shows

The British tax system would claim that this is not backwards legislation as the tax assessment is done at $t_2 + 1$. However, for the legislation not to be backwards one should adopt the meaning of "ship" as it was at $[t_1, t_2]$ for the purpose of tax paid at $t_2 + 1$ on the period $[t_1, t_2]$.

We understand that countries like Austria, never legislate backwards. It is taboo! So the new definition of "ship" will be applied only to employment after t_2!

Let us summarise what we need for our logic. We need to allow for time dependent predicates of legal and social nature generated by actions. We write them as follows:

- $t \vDash P(x_1, \ldots, x_n)$ if action $\mathbf{a}$ is taken by $x_1, \ldots, x_n$ at time t satisfying the pre-conditions $\mathbb{C}_{\mathbf{a}}$.

- $t \vDash \neg P(x_1, \ldots, x_n)$ if either no action $\mathbf{a}$ was taken in the past or a cancelling action $\mathbf{b}$ was taken at t, with preconditions $\mathbb{C}_{\mathbf{b}}$.

It is with this sort of predicates we want to present our logic **TTL**.

EXAMPLE 26. Two security agents meet in a bar having a beer and discussing their profession. Say Microsoft chief security officer **m** and Google chief security officer **g**. **m** boasts to **g** that his methods are impregnable. **g** admits **m** is good but not perfect. **m** challenges **g**. He says:

> I have a laptop in my office which is security protected. I shall clear the disk drive and leave it on the internet. At 18.00 hours it will be security protected and I shall call you and give you this laptop to be immediately yours on the condition that you break into it within 30 minutes.

m cautions **g** that he had better not trip any alarms because it is illegal to hack into the system.[58]

We have here three periods of time

1. Before 18.00
2. From 18.00 to 18.30.
3. After 18.30

During period 2 it is not clear who owns the laptop.

After 3, the situation clarifies. There are two approaches of how to model the situation.

1. *The Fisher approach*[59]
   From 18.00–18.30 we have two parallel histories. One in which the laptop belongs to **m** and **g** is unable to hack into it and second in which the laptop belongs to **g** (from18.00) and **g** was able to hack into it. At 18.30 we know which history is real. At period 3, after 18.30, there is only one history.

   Note that according to Fisher, no crime has been committed by **g**. Even if he managed to hack into the laptop, this made the laptop his from 18.00 and there is no crime to hack into one's own laptop.

   Furthermore, even if **m** dies between 18.00 and 18.30, the deal is on — nothing changes!

   Also if **m** changes his mind at 18.15, he cannot cancel the deal. Ownership of the laptop has already been (conditionally) transferred at 18.00!

2. *The Shkop approach*[60]
   This view says that the deal is completed and actually executed at 18.30. Hence

[58] According to British law **m**, by making the offer to **g** is already giving him permission to hack into the laptop.

We can change the example a bit. **m** sells the laptop to **a**. **a** makes the condition that if anyone hacks into it between 18.00–18.30 then the deal is off. Now **m** gives the laptop to **g** under the condition that **g** hacks into it between 18.00–18.30. Now **g** would commit a crime.

[59] Rabbi Shlomo Fisher, 1932–.

[60] Rabbi Shimon Shkop, 1869–1939.

(a) **g** commits a crime in hacking into the laptop because at the time of hacking the laptop was not yet his.

(b) If **m** dies at 18.15 or changes his mind and cancels, then the deal is off as he (**m**) is not there to execute the deal at 18.30.

How do we model Shkop's view? We need dual time t and τ:

$$\begin{array}{rcl} t & = & 18.00, \ldots, 18.15, \ldots, 18.30, \ldots, 19.00 \\ \tau & = & \text{simulated time in minutes.} \\ & & \text{We start at } \tau = 0 \text{ at } t = 18.00, \text{continue to } \tau = 30 \text{ at} \\ & & t = 18.30 \text{ and immediately go back to } t = 18.00,^{61} \\ & & \text{At } \tau = 30 \ \mathbf{m} \text{ and } \mathbf{g} \text{ complete the deal, then carry on to} \\ & & t = 18.30, \tau = 60 \text{ and continue to eternity with } t \text{ and } \tau. \end{array}$$

Think of it as that both **m** and **g** jumping instantly back from time $t = 18.30$ to $t = 18.00$ using a time machine and completing the deal. Their personal time is τ. They live through history again and are back at $t = 18.30$ with their personal time τ being 60 minutes.

With τ, we remember that crime was committed at $\tau = 30$. If **m** dies at 18.15 then he cannot go back at 18.30 ($\tau = 30$) to the beginning ($t = 18.00, \tau = 30$), to complete the deal.

REMARK 27 (Semantic discussion of the Fisher approach). Let us appreciate the difficulties in modelling the Fisher approach. The Fisher view gives rise to a temporal history without memory. If you go with time to infinity then there is only one linear past without any memory that it could have been otherwise. It cannot be modelled by branching time with one infinite branch being the real history because it allows for memory of alternatives.

EXAMPLE 28 (Iterated conditions). Once we allow conditional legal actions of the form **a** at t conditional on **b** at $t + s$, we should be able to iterate. Figure 5 shows such a case.

The story is as follows: John gives Mary a pen to be her property immediately at time t on the condition that she buys some shares at time $t + s$. At time $t+s$ Mary approaches Terry who has shares and he is willing to sell Mary his shares to be hers immediately on the condition that Mary does his garden on $t + s + r$.

The first comment we make is that we consider we have two actions here:

1. Action **a** starting at time $x = t$ with condition at time $x + s$.

2. Action **b** starting at time $y = t + s$ with condition at time $y + r$.

[61] At $\tau = 30$ there is a discontinuous jump from $t = 18.30$ back to $t = 18.00$.

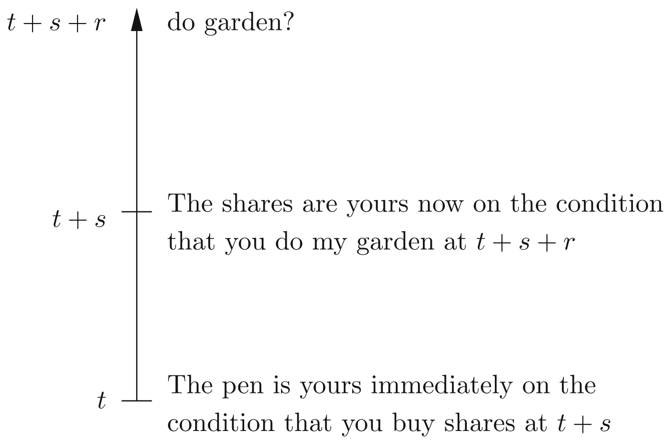


Figure 5.

According to Fisher we have two conditional actions which are chained by making basically $y = x + s$.

According to Shkop, we also have two actions which are chained by making $y = x + s$. However according to Shkop the actions retain their identity as distinct actions in the sense that each action has its own τ. So action **a** has τ and action **b** has τ'.

According to Fisher, we have two parallel histories. See Figures 6, 7.

The real history is decided on time $t + s + r$.

According to Shkop, Figure 5 turns into Figure 8 as follows.

We start with action **a** at t and $\tau = 0$. At $t + s$ we have $\tau = s$. We want to jump back using our time machine and be again at t but with $\tau = s$. We ask ourselves: where are the shares? Do we have them at time $t + s$ (with $\tau = s$)? Can we jump back? Is the matter of the shares decided at $t + s$ so that we can jump back and conclude the deal if Mary bought the shares or cancel it if Mary does not have the shares? The answer is that we don't know yet, it depends on action **b**. Well, we have the shares but on the condition of Mary doing the garden at $t + s + r$.

OK then. We cannot jump back with τ at $\tau = s$, because we have to wait for action **b** to play itself out with its own τ'.

So action **a** does not jump back to t, action **a** has to wait for action **b**. So both action **a** and action **b** proceed together to time $t + s + r$.

So now $\tau = s+r$ and $\tau' = r$ and both actions are sitting at time $t+s+r$. Action **a** is waiting for action **b** to jump with its τ' back to time $t + r$ and decide the matter of whether Mary buying the shares is successful. So τ' jumps back to time $t + r$ and decides the matter of Mary's owning the shares (depending on her doing the garden at time $t + s + r$) and then τ'

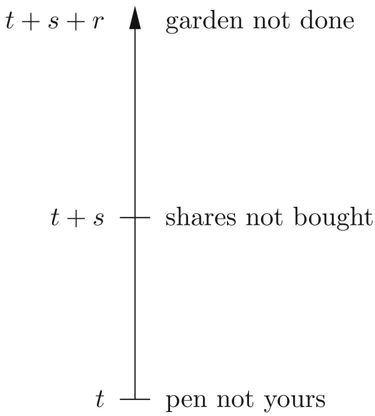


Figure 6.

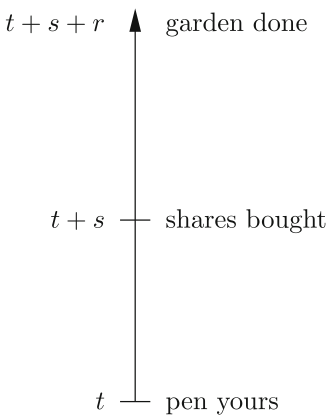


Figure 7.

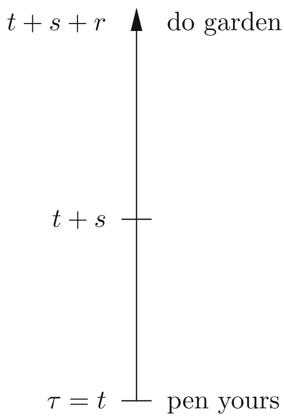


Figure 8.

goes back to time $t+s+r$. By this time $\tau' = 2r$. While τ' was jumping back τ was sitting at $t+s+r$, with $\tau = s+r$, waiting for an answer from **b**. The changes in τ' have nothing to do with τ. It is an internal action **b** simulation.

Once action **a** gets an answer from **b** it jumps back to time t, decides whether the pen belongs to Mary and proceeds back to time $t+s$. At this time $\tau = (s+r)+s$. Now it is known whether the condition **b** holds so τ jumps back to t and proceeds straight back to $t+s+r$. By this time $\tau = (s+r)+(2s+r) = 3s+2r$.

Notice that in this case action **a** is equivalent to a new action $\mathbf{a}'$ comprised of the giving of pen at time t on the condition of buying shares at $t+s$ with a final decision time of doing the garden on $t+s+r$.

The following is the scenario for action $\mathbf{a}'$.

The τ count starts with $\tau = 0$ at at t! See Figure 8.

The situation clarifies and the deal is executed on time $t+s+r$ with $\tau = s+r$. At that time we go back to t to execute the deal and $\tau = s+r$ carries on through history again back to time $t+s$ where another jump back is done to time t and then we go straight to time $t+s+r$ with $\tau = (s+r)+(2s+r) = 3s+2r$.

EXAMPLE 29 (Logical loop). Let us analyse now case 4 of example 24.

> Time $t = 0$: a gives divorce to b on the condition that she never marries c
> Time $t = s$: [b has married c before time s] or [b has not married c before time s]

Analysis according to Fisher:

We have two parallel histories beginning at $t = 0$ and ending at $t = s$. At $t = s$ a decision is made as to which history is real.

History 1:
[At $t = 0$, b is divorced from a] and [at $t = s$, b has not married c before time s]
History 2:
[At $t = 0$, b is not divorced from a] and [at $t = s$, b has married c before time s].

Obviously History 2 is legally inconsistent and therefore History 1 prevails.

Analysis according to Shkop:
Start at $t = 0$ and $\tau = 0$. Continue to $t = s$ and $\tau = s$.

Case 1:
If b has not married c up to s then carry on with t and τ to $s+1$ and repeat the case analysis.
Case 2:
If b has indeed married c then assume s is the first time this is done.
Jump back to $t = 0$ and $\tau = s$ and cancel the divorce.
Continue forward from this point (i.e. $t = 0$ and $\tau = s$) and reach any $t = r$ and $\tau = s + r$ for any r. b can never marry anybody at $\tau = s + r$ since she is already married to a since $\tau = s$.

EXAMPLE 30 (Two actions). Let us have two actions. One giving the laptop at $t = 18.00$ and the other giving the pen also at $t = 18.00$. What do we do?

We need two τ counts. One for the laptop, τ_1 and one for the pen, τ_2. In fact, during a normal history with many actions and many chains, we have as many τs counting simulated time.

REMARK 31 (Analysis of chains). We now want to analyse Example 28 and prepare ourselves for Example 33.

Let us start with action $\mathbf{b}$ of Example 28. This action starts at an abstract time y (which was instantiated as $y = t + s$) trying to make the predicate $P_{\mathbf{b}}$ = "Mary owns shares" true at y.

The truth value of the predicate was not clarified until time $y + r$. At this time the final predicate $Q_{\mathbf{b}}$ = "Mary doing the garden" was the one whose truth value clarified the status of the starting predicate $P_{\mathbf{b}}$.

Taken in the abstract, the relevant parameters of action $\mathbf{b}$ are as follows:

1. Starting time y
2. Predicate involved is $P_{\mathbf{b}}$
3. Stretch of the action, namely the duration until the predicate $Q_{\mathbf{b}}$ is to be determined, is r (i.e. it goes from y to $y + r$)

4. The final predicate which clarifies the state of the predicate $Q'_{\mathbf{b}} = Q_{\mathbf{b}}$, at the same time $y + r$.

Let us now do a similar analysis for action **a**.

1. Starting time is x
2. Predicate involved is $P_{\mathbf{a}}$, ("Pen belongs to Mary at time x")
3. Stretch is s, with predicate $Q_{\mathbf{a}}$ at time $x + s$
4. The final predicate which clarifies the status of $Q_{\mathbf{a}}$ is $Q'_{\mathbf{a}} = Q_{\mathbf{a}}$ also at time $x + s$.

How do we make a chain of these two abstract actions? We equate the final predicate of **a** with the initial predicate of **b** and say at what time. In example 28 we did the following:

1. Let $Q'_{\mathbf{a}} = P_{\mathbf{b}}$
2. For the time let Equation (x, y) be: $y = x + s$

This chaining resulted in a new action, which we called $\mathbf{a}'$:

1. Starting time is x
2. $P_{\mathbf{a}'} = P_{\mathbf{a}}$
3. Stretch is s with $Q_{\mathbf{a}} = P_{\mathbf{b}}$
4. Final predicate $Q'_{\mathbf{a}}$ is $Q_{\mathbf{b}}$ at time $x + s + r$.

Note that we could have chosen a different equation for the combination of **a** and **b**, we could have chosen $y = x + s - 1$. In this case we would have got a new action, say $\mathbf{a}''$ with stretch $s + r - 1$.

In practice, when combining actions such as **a** and **b**, one does not write any equation between x and y. When x and y are realised in a real time model, they get specific time values, and the equation is determined automatically. Example 33 below is such an example.

The exact formal definitions of action combination is worked out in Section 6.

REMARK 32. Note that we are dealing here with a single condition $Q_{\mathbf{a}}$ for the action **a**. In other words the conditional is of the form:

- $P_{\mathbf{a}}$ now at time t, on the condition that $Q_{\mathbf{a}}$ later at time $t + s$.

For example

- The pen is yours immediately now at time t, on the condition that you buy shares at $t + s$.

We have only one atomic condition and no more. So we are not addressing multiple conditions of the form:

- $P_{\mathbf{a}}$ now at time t, on the condition that for $i = 1, \ldots, k$ we have $Q_(i, \mathbf{a})$ holds later at time $t + s_i$.

For example we are not dealing with:

- The pen is yours immediately now at time t, on the condition that you buy shares at $t + s$ and put your computer on Ebay at $t + r$.

There is no technical difficulty in addressing multiple conditions, it is just that such examples do not appear in the Talmud in this form.

The Talmud can have conditions of the form:

- The pen is yours immediately now at time t, on the condition that you DO Not sell your shares BEFORE time $t + s$.

This has the formal form:

- $\neg P_{\mathbf{a}}$ now at time t, on the condition that $Q_{\mathbf{a}}$ holds at a time r such that $t < r < t + s$

or equivalently

- P_a now at time t, on the condition that $Q_{\mathbf{a}}$ holds at all times r such that $t < r < t + s$.

We may have some difficulties with chaining such conditions. Obviously we have no problems chaining if maintaining that $Q_{\mathbf{a}}$ holds at all times r such that $t < r < t + s$ is enabled by some condition $\mathbf{b}$ executed after time s. However what if for each $r, t < r < s$ we need to promise a separate condition $\mathbf{b}(r)$ to ensure that $Q_{\mathbf{a}}$ holds at r?

This would fall under item (2) of Definition 38 below.

EXAMPLE 33 (Cross chain dependency). We start with the chain action of Figure 7. This is the chain action discussed already in Example 28 and Remark 31. We now want to chain into it a new action $\mathbf{c}$.

The laptop is yours at $t - 1$ provided the pen is yours at $t + s + \frac{r}{2}$ (of Figure 7).

We have two actions here to be synchronised. See Figure 9.

We start counting $\tau_1 = 0$ at $t - 1$. We get to time $t + s + \frac{r}{2}$ with $\tau_1 = 1 + s + \frac{r}{2}$ and ask " Is the pen yours?".

Well, at this time there is the τ_2 counting of the pen. τ_2 counting at $t + s + \frac{r}{2}$ is at $\tau_2 = s + \frac{r}{2}$. We have to wait another $\frac{r}{2}$ for τ_2 to get to $s + r$ and τ_1 to get to $1 + s + r$. The real time is now $t + s + r$. Then τ_2 has to go back to t to complete the pen deal and advance back to time $t + s$, double back to t and then proceed to $t + s + r$. This takes $\tau_2 = 3s + 2r$ minutes.

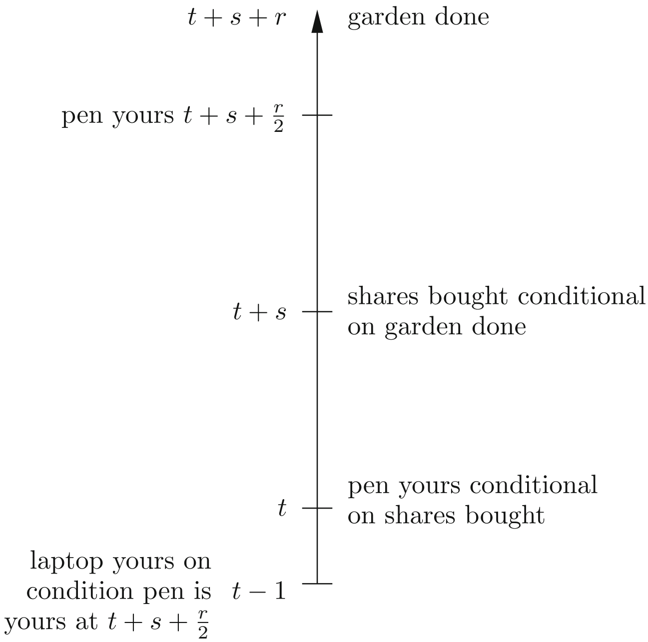


Figure 9. Right hand side = pen ownership conditions; left hand side = laptop ownership conditions

Note that τ_1 does not change, it does not care what τ_2 does. So τ_1 is equal $(1 + s + r)$. When τ_2 reaches time $t + s + r$ it "informs" τ_1 that the pen deal is done. Now τ_1 jumps back to $t - 1$ to complete his deal. The jumping is from real time point $t + s + r$. τ_1 advances another simulated time from $t - 1$ $t + s + \frac{r}{2}$ where the deal is supposed to be done and then jump back to t to clinch the deal and then proceed straight to $t + s + r$ because the real time $t + s + r$ is where τ_1 is. He can now confirm the deal is done. This brings us to

$$\begin{aligned} \tau_1 &= (1 + s + r) + (1 + s + \tfrac{r}{2}) + (1 + s + r) \\ &= 3 + 3s + \tfrac{5}{2}r. \end{aligned}$$

Note that $\tau_2 = 3s + 2r$ as calculated in Example 28.[62]

From $t + s + r$ real time τ_1 and τ_2 continue to tick.

EXAMPLE 34 (Contrary to duties in the Talmud). These have been analysed in [5; 13]. Some of them are temporal, what we called Type CTD III.

You should not steal, and if you did steal, you have an obligation to return or pay for the stolen object. If you do return the stolen object, the violation is cancelled retrospectively. This is why the Talmud does not recommend immediate punishment for stealing because the action might be cancelled retrospectively in the future by returning the object.

We make two relevant comments here.

1. In the case of stealing, Rabbi Shkop agrees with the Fisher model. So his Shkop model applies only to conditional actions and not to Contrary to Duties.

2. Since the stealing can be cancelled retrospectively in the future, one might adopt the view of habitually stealing objects and then cancelling the action by returning the objects stolen, and so he has no sin, but lots of "temporarily stolen" objects, which he returns again and again.

   This is reminiscent of the case where a person has no income and no tax to pay, because he only borrows the money at the beginning of a tax year to return it at the end of the tax year, only to immediately borrow it again at the beginning of the next tax year.

   There is an extensive discussion in the Talmud of how to deal with such cases.

[62]Note that what we call τ_2 here is called τ in Example 28, it is the τ of action $\mathbf{a}'$ at that example.

3.9 Talmudic temporal logic

We begin with the definition of Talmudic action system. In order to present it properly, let us start with existing simple action systems of artificial intelligence.

In ordinary artificial intelligence an action for a certain language $\mathbf{L}$ has the form $\mathbf{a} = (\alpha_{\mathbf{a}}, \beta_{\mathbf{a}})$, where $\alpha_{\mathbf{a}}$ is the precondition of the action and $\beta_{\mathbf{a}}$ is the post condition of the action, $\alpha_{\mathbf{a}}$ and $\beta_{\mathbf{a}}$ are in the language $\mathbf{L}$.

We can be more specific about this form. Let t be a moment of time (in which the action takes place) and let $x_1, \ldots, x_n$ be the individuals involved in the action. We can write $\alpha(x_1, \ldots, x_n), \beta(x_1, \ldots, x_n)$ as the preconditions and post condition of the action and write $\mathbf{a}(x_1, \ldots, x_n)$ to indicate that the action $\mathbf{a}$ involves the individuals $x_1, \ldots, x_n$.

We can now write $\mathbf{Exec}(t, \mathbf{a})$ to indicate that the action $\mathbf{a}$ was executed at time t.

In ordinary AI there are no backward causal actions and so all we have is the above. We can specialise it a bit, like asking that $\beta_{\mathbf{a}}$ be an atomic predicate.

Note that the preconditions and post-conditions are not time dependent.

DEFINITION 35 (Classical action temporal logic).

1. The language of classical action temporal logic has the following components:

   1.1. Variables and constants for time points $t_1, t_2, \ldots, \mathbf{t}_1, \mathbf{t}_2, \ldots$

   1.2. Variables and constants for domain elements, $x_1, x_2, \ldots, \mathbf{d}_1, \mathbf{d}_2, \ldots$.

   1.3. The classical connectives and quantifiers for two sorted logic.

   1.4. A set of n-place action names with domain variables or constants of the form $\mathbf{a}(x_1, \ldots, x_n)$

   1.5. A unary existence predicate $E(x), x$ domain variable.

   1.6. An execution predicate of the form $\mathbf{Exec}(t, \mathbf{a}(x_1, \ldots, x_n))$.

   1.7. The earlier–later predicate $t < s$ for time variables.

   1.8. n-place atomic time + domain predicates $P(t, x_1, \ldots, x_n)$ with t time variable and x_i domain variables.

   1.9. We define traditionally the usual notion of a time domain formula $\varphi(t_1, \ldots, t_k, x_1, \ldots, x_n)$ using the atomic predicates in 1.5, 1.6, 1.7, 1.8 and the connectives and quantifiers in 1.3.

   1.10. We associate with each action $\mathbf{a}(x_1, \ldots, x_n)$ two formulas $\alpha_{\mathbf{a}}(t, x_1, \ldots, x_n)$ and $\beta_{\mathbf{a}}(t, x_1, \ldots, x_n)$ as defined in 1.9. We assume $\beta_{\mathbf{a}}(t, x_1, \ldots, x_n)$ is atomic as defined in 1.8. The variables in $\alpha_{\mathbf{a}}$ and $\beta_{\mathbf{a}}$ are as indicated. We call $\alpha_{\mathbf{a}}$ the precondition for $\mathbf{Exec}(t, \mathbf{a}(x_1, \ldots, x_n))$ and $\beta_{\mathbf{a}}$ the post-condition. Note that $\alpha_{\mathbf{a}}, \beta_{\mathbf{a}}$ and $\mathbf{Exec}$ have the same variables $(t, x_1, \ldots, x_n)$.

2. A model $\mathbf{m}$ has the form $\mathbf{m} = (T, <, \mathbb{A}, D, h)$ where $(T, <)$ is a flow of time, say linear flow, $\mathbb{A}$ is the set of actions and D is a domain of elements. $T, \mathbb{A}$ and D are pairwise disjoint. h is an assignment giving to each n-place predicate P a subset $h(P) \subseteq T \times D^n$. For each n-place action $\mathbf{a}(x_1, \ldots, x_n)$ and each $d_1, \ldots, d_n \in D$ and each $t \in T$ we have $h(\mathbf{Exec}(t, \mathbf{a}(d_1, \ldots, d_n)) \in \{0, 1\}$.

   The truth value of a wff $\varphi(x_1, \ldots, x_n)$ is defined by induction in the traditional manner.

$$\begin{array}{lll} \mathbf{m} & \models & P(t, x_1, \ldots, x_n) \text{ iff } (t, x_1, \ldots, x_n) \in h(P) \\ \mathbf{m} & \models & \mathbf{Exec}(t, \mathbf{a}(x_1, \ldots, x_n)) \text{ iff } h(\mathbf{Exec}(t, \mathbf{a}(x_1, \ldots, x_n))) = 1 \\ \mathbf{m} & \models & \text{the connectives and quantifiers in the traditional manner} \end{array}$$

   We require some integrity constraints to hold, for example

   - $\mathbf{m} \models \mathbf{Exec}(t, \mathbf{a}(x_1, \ldots, x_n)) \rightarrow \alpha_{\mathbf{a}}(t, x_1, \ldots, x_n) \wedge \beta_{\mathbf{a}}(t, x_1, \ldots, x_n)$
   - $\mathbf{m} \models \beta_{\mathbf{a}}(t, x_1, \ldots, x_n)$ iff $\exists s \leq t[\mathbf{m} \models \mathbf{Exec}(s, \mathbf{a}(x_1, \ldots, x_n))$ and for all u, $s \leq u \leq t$ and all $\mathbf{b} \in \mathbb{A}$ such that $\beta_{\mathbf{b}} = \neg\beta_{\mathbf{a}}$ we have $\mathbf{m} \not\models \mathbf{Exec}(u, \mathbf{b}(x_1, \ldots, x_n))$
   - Note that the pre-conditions of actions do not change with time. So, for example, if a foreign language is requied for a PhD it is always a requirement.

EXAMPLE 36 (Conditional actions). Now that we have a more exact formalism for actions, let us reconsider the examples of Section 3.8. Consider Example 26. We have to specify more precisely the pre-conditions and post-conditions of each action.

1. *Action x gives laptop ownership to y.*
   Preconditions:

   - x owns the laptop
   - y is allowed to own the laptop
   - a document is written transferring ownership
   - x and y exist and sign document

Postconditions

- x does not own the laptop
- y owns the laptop
- document exists

2. *Action y hacks into x's laptop undetected.*
Preconditions:

   - laptop exists
   - y exists

   Postconditions:

   - y logged onto laptop
   - no alarms triggered.

In the Shkop model of the conditional of Example 26, we said that at time 18.30 both agents **m** and **g** go back in time to 18.00 and conclude the deal. The question is which of the pre-conditions of the action of "give laptop ownership" we require to hold at 18.30?

Obviously **m** and **g** need to exist at 18.30.[63] Do we also require that the document exists? What if at 18.15 the document was destroyed? Well, this depends on the legal system. In the Talmud, for example, to have an effective divorce agreement, the document must exist! Another question is do we need the original document, or can a new one be drawn at 18.15 if the original one was destroyed? Obviously we need to specify, when we make a conditional of the form

Action **a** at t if Action **b** at $t+s$

which pre-conditions of Action **a** should hold at $t+s$ before we "jump" back (in the Shkop model).

For this reason we present the pre-conditions of any action **a** as a pair of formulas

$$\mathbb{C}_{\mathbf{a}} = (\alpha_{\mathbf{a}}, \gamma_{\mathbf{a}})$$

Both have to hold in order for the action to be executed. However, $\gamma_{\mathbf{a}}$ is the one that passes on to the future if we make **a** conditional on some future **b**. So for example in the laptop case,

[63] Note that in the case that one of them dies exactly at 18.30, this still counts as "existing" at 18.30, for the purpose of the model. This follows from Talmudic rulings in such cases. So, according to Shkop, the Talmud requires them to exist at all moments of time up to but not necessarily including the end time 18.30.

$\alpha_{\mathbf{a}} =$ document signed and exists, x owns the laptop and y allowed to own laptop

and

$\gamma_{\mathbf{a}} =$ x and y exist.

REMARK 37.

1. It stands to reason to say that all pre-conditions of actions **a** are always pure and unconditional. Otherwise we put them as conditionals for **a** itself. So to make it clear, if the ownership of the laptop by **m** is itself conditional on **c**, then **m** can give the laptop to **g** only if he makes the action conditional on the condition **c** as well as any other conditions he may wish to add.

2. We have seen in Examples 23 and 26 that preconditions of actions can be ignored and the results of the illegal action can be used for backwards causality. Also not all the postconditions of the action need to be recorded but only those relevant tothe backwards causality. Therefore the facts of interest to our models are

   - whether an action can be executed, legally or not
   - what post conditions are relevant
   - what preconditions can, if not satisfied, block the execution of the action.

3. We can also assume that the "condition" is a state caused by some action. It could be a state of "being married to c" caused by the action of conducting the marriage ceremony, in which case if the woman is already married to a with a different from c, and so the action has no consequence. It could also be the state of "having executed a marrige ceremony with c" in which case the state is achieved by the action, even if the woman is already married to a.

DEFINITION 38 (Linear chain of conditional actions, preliminary version).

1. *Level* 0 *(no condition) actions*
   These have the form $\mathbf{a}(\mathbb{C}_{\mathbf{a}}, \beta_{\mathbf{a}})$ where $\mathbb{C}_{\mathbf{a}} = (\alpha_{\mathbf{a}}, \gamma_{\mathbf{a}})$ are the pre-conditions and $\beta_{\mathbf{a}}$ is the post-condition. We assume that if **a** is used in conditionalised form then $\alpha_{\mathbf{a}}, \gamma_{\mathbf{a}}$ will be required to hold at different times, as discussed in (2) below.
   Level (1) *actions*
   These have the form
   a at t if **b** at all u such that $t < u \leq s$ and $\varphi(t, s, u)$,
   where **a** and **b** are level (0) actions and where t, s are temporal constants $t < s$ and φ a temporal statement about the interval $[t, s]$. We

allow for s to be infinity $s = \infty$. For example $\varphi(t, s, u) \equiv (s = u)$ or $\varphi(t, s, u) \equiv (t < u \leq s)$ or $\varphi(t, s, u) \equiv (t < u)$.
Level $(n + 1)$ *actions*
These have the form
a at t if **b** at all u such that $\varphi(t, s, u)$
where **a** is a level one action and **b** is level $(n + 1)$ action.

The above defines simple linear chains.

2. *General inductive clause*
The general definition is as follows:

Let **a** and **b** be any conditional action already defined, then:

a at t if **b** at all u such that $t < u \leq s$ and $\varphi(t, s, u)$
is also an action where $t < s$

3.10 Conclusion

We introduced in this section the Talmudic Temporal Logic, capable of modelling the Talmudic examples. The logic was motivated and introduced semantically. We are not going to develop its formal properties, proof theory, completeness, its relation to other logics, etc., etc. This is the subject for another, pure logic paper and is not essential for modelling the Talmud.

Note however that Talmudic reasoning does “export” to general logic new ideas about temporal causality.

Some of our examples in this section dealt with entities defined using the future. These are dealt with in the next section, see also [93]. We shall see that, again, the Talmud exports to general logic a new type of public announcement logic with quantum superposition semantics.

4 FUTURE ORIENTED DETERMINATION OF ENTITIES IN TALMUDIC LOGIC

Preliminaries

Ordinary dynamic action logics deal with states and actions upon states. The actions can be deterministic or non-deterministic, but it is always assumed that the possible results of the actions are clear cut.

Talmudic logic deals with actions (usually legally meaningful actions which can change the legal status of an entity) which may be not clear cut and need clarifications.

The clarification is modelled by public announcement which comes at a later time after the action has taken place.

The model is further complicated by the need to know what is the status of formulas at a time before the results of the action is clarified, as we do not know at which state we are in. Talmudic logic treats such states much like the quantum superposition of states and when clarification is available we get a collapse onto a pure state.

The Talmudic lack of clarity of actions arises from applying an action to entities defined using the future, like the statement of a dying man on his death bed:

> **Let the man who will win the jackpot in the lottery next week be the sole heir in my will now**

We need to wait a week for the situation to clarify.

There is also the problem of legal backwards causality, as this man, if indeed he exists, unaware of his possible good fortune, may have himself meanwhile donated all his property to a charity. Does his donation include this unknown inheritance?

This paper will offer a model and a logic which can represent faithfully the Talmudic reasoning in these matters.

We shall also see that we get new types of public announcement logics and (quantum-like) action logics.

4.1 Introduction

The Talmudic logic we are going to construct is comprised of several components, some known to us already and some are new. The Talmudic system is then used to model certain aspects of reasoning in the Talmud.

The fragments of the logics we are going to use to combine and construct our final Talmudic logic system are as follows:

1. Some aspects of modal **K** action logic.
2. Some aspects of public announcement logic.
3. Some aspects of the logic of time.
4. Some aspects of (quantum-like) superposition and collapse.

We begin by explaining the effects of these components.[64]

Imagine a modal S5 logic of the form (S, t),[65] where S is the set of possible worlds and $t \in S$ is the actual world. Suppose we perform an action $\mathbf{a}$ in the world t which moves us from the world t to the world s where s is the world where the post condition of the action holds. Schematically we have Figure 10.

[64] For the convenience of the reader, who might not be familiar with one or more of these components, we are including a short exposition in the Appendix.

[65] We use the letter "t" for the actual world, even though it is usually reserved for time points. In our intended models (after actions are clarified) time is linear and discrete and time points are the results of the application of linear sequences of actions, and so the worlds are the times. See Definition 53.

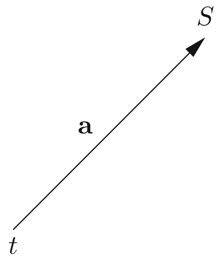


Figure 10. Change of state following an action

The model after the action is (S, s). To be specific, consider a model $\mathbf{m}_t$ with three element domain $D = \{a, b, c\}$ and a unary predicate $\lambda x P(x)$.[66] Assume that

$$\mathbf{m}_t \vDash \neg P(a) \wedge \neg P(b) \wedge \neg P(c).$$

The action $\mathbf{a}$ is to make P hold for exactly one of the elements.

So if we describe the action $\mathbf{a}$ as

$$\text{execute “}\exists! x P(x)\text{”}$$

then it is non-deterministic and can have three outcomes:

$$\begin{array}{l} s_1 \vDash P(a) \wedge \neg P(b) \wedge \neg P(c) \\ s_2 \vDash \neg P(a) \wedge P(b) \wedge \neg P(c) \\ s_3 \vDash \neg P(a) \wedge \neg P(b) \wedge P(c) \end{array}$$

If we must have perfect clarity we must execute one of the options above, i.e.

either execute “$P(a)$”
or execute “$P(b)$”
or execute “$P(c)$”

So far we have a very simple action logic, where actions $\mathbf{a}$ performed in one world t take us to a clear cut unique world $\mathbf{s}$.

Let us now complicate the situation. Suppose the action was done in such a way that it is not clear whether the result is world s_1 or s_2 or s_3. We represent this situation in Figure 11.[67]

[66]The reader should note that we are dealing with finite domains and therefore the logics involved are propositional, not predicate logics.

The universal quantifier can be rewritten as a conjunction over all elements of the domain and the existential quantifier can be rewritten as a disjunction over all elements of the domain.

So because our work looks like a first-order logic, it makes it easy to express superposition of elements but at the same time since the logic is really propositional, we need not worry about the well known difficulties in the treatment of first-order

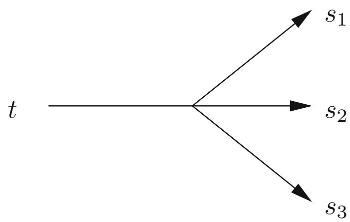


Figure 11. Lack of clarity

We are not necessarily saying that the action is non-deterministic and allows for several possible outcomes (with some probability) as they do in Agent Theory and Dynamic Logic. We are thinking that perhaps it wasn't clear exactly what happened or that the action was interrupted or any other reason for us to have to wait until the matter clarifies.

This is why we use special notation, see Figure 12.

So we expect a clarification, a public announcement, telling us where we are.

Figure 13 describes the situation all in terms of public announcement.

In Figure 13, the arrows are schematic, they do not take time. In the Talmud, the models are temporal and the arrows take time. Figure 14 shows what happens in time in the Talmud.

To make the example real, suppose an American billionaire is an admirer of three football players, a, b and c. He writes a will at time 1 leaving one of them exactly as his sole heir. In ten days time they are all going to play in a football match. The one who scores most goals is the heir. If two or three score the same number of goals a lottery is used to choose the heir. There is some lack of clarity in the will which clarifies at time 10.

In this case our predicate is $P(x) =$ "x is the sole heir".

So the clarification takes time and it is at this point that the temporal aspect comes in.

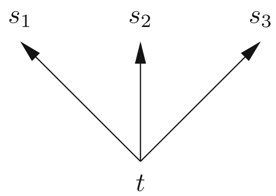
We need to address the following problems.

1. Analyse the nature of the (Talmudic) action which can give rise to the lack of clarity at time 1.

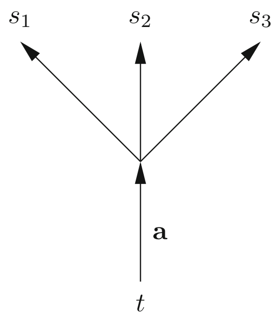
---

modal/epistemic/deontic/temporal logics.

[67]Let us be more specific here. Suppose I ask my agent, (from Mrs Renton matrimonial services) to go and arrange for my engagement to one of the three candidates $\{a, b, c\}$. I leave it to my agent to decide who is most suitable. The interview is scheduled for Monday. In this case one of them is engaged to me on Monday, though I may not know which one it is until the following Friday. Another possible scenario is that I give my agent a ring and ask him to arrange for my engagement to one of $\{a, b, c\}$ on Monday. The agent gives the ring to all three of them and says he will inform them later which one he chooses. In this case on Monday one of them is engaged but for each one of $\{a, b, c\}$ we cannot say she is engaged.



Non-deterministic action **a**



Deterministic action **a** in need of clarification

Figure 12. Non-deterministic action

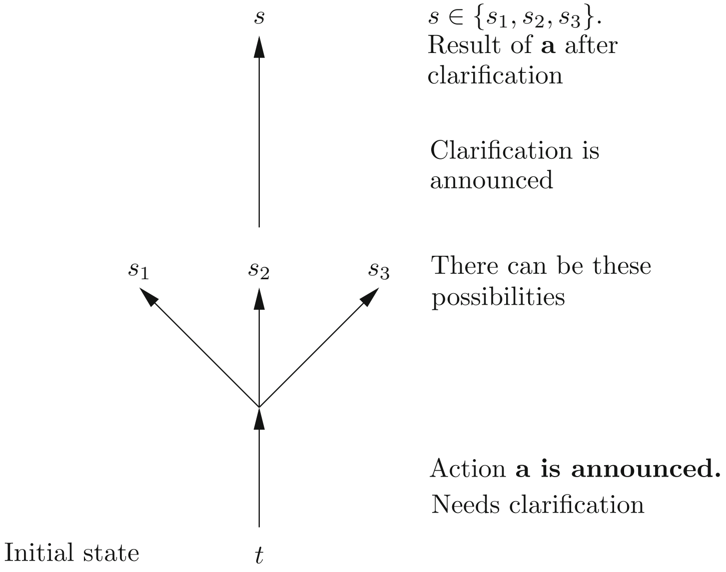


Figure 13. Public announcement clarification

1.1 Identify parameters in the action which cause the lack of clarity.

1.2 Determine what are the possible results of this lack of clarity.

By possible results we mean that if action $\mathbf{a}$ is applied to model $\mathbf{m}_t$ and action $\mathbf{a}$ has component α causing lack of clarity, what is the list of possible models $\mathbf{m}^{\alpha}_{s_1}, \ldots, \mathbf{m}^{\alpha}_{s_n}$ which can result.

We await clarificiation as to which $\mathbf{m}^{\alpha}_{s_i}$ does result. Note that in Figure 14 the possible results of action $\mathbf{a}$ on $\mathbf{m}_0$ are $\mathbf{m}_1$ or $\mathbf{m}_2$ or $\mathbf{m}_3$. These are all classical models. Note that we cannot adopt the view that we are, at time 1, after the execution of the action, either at model $\mathbf{m}_1$ or at model $\mathbf{m}_2$ or at model $\mathbf{m}_3$. To see this, think of predicate P to mean being infected. So as a result of the action, either a or b or c were infected. At time 2, we would like to put all infected people into isolation. Common sense says we need to isolate all three elements $\{a, b, c\}$. This action cannot be justified when applied to the situation in each of the models $\mathbf{m}_1$, $\mathbf{m}_2$ or $\mathbf{m}_3$. Clearly the result of the lack of clarity puts us in a new model $\mathbf{m}_4$. The Talmud allows (because of the lack of clarity) additional models which are some sort of (quantum-like) superposition models,[68] which are different from

[68]It should be noted that the quantum-like superposition is itself the model here — and not a logicized form of quantum mechanics as attempted in some forms of "quantum

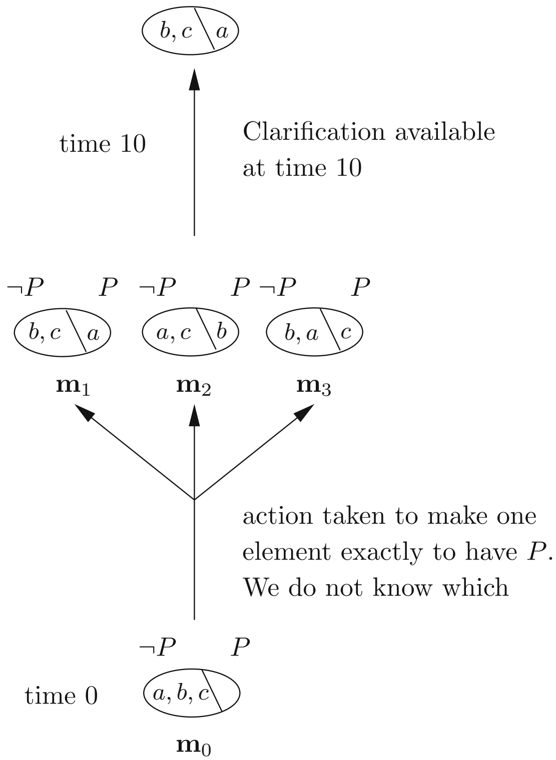


Figure 14. Time action in the Talmud

the classical models. So we will have to define some new Talmudic models of our logic.

1.3 We ask: can the situation be clarified? Do we treat these "unclear" actions as a new type of actions, and a new type of action logic?

Furthermore, if we have new superposition models, how does the old action **a** apply to the new model?

2. Once we have clarity at time 10, do we apply it backwards from time

logic". For a survey of attempts at conveying the logical form of a quantum-mechanical state (physical, mathematical) or logicize vagueness and fuzzyness, [37]. The Superposition principle is what makes the Hilbert spaces and their subsets into (quantum mechanically) more than classical [37, p. 207]. Quantum mechanics deals with different sorts of fuzziness, the "sharp kind" —– where "events are sharp, while all semantic uncertainties are due to the logical incompleteness of the individual concepts, that correspond to pure states of quantum objects." (such as determination of Hamlet's height — [37, p. 245] and an "unsharp" vagueness where the predicate itself is not well defined (the honor of Brutus). See also [99] for general reading.

1?

If not, what is the status (where are we? at which world? at which model?) at times 2 to 9?

We can see that we need to develop at least 3 new logics, maybe 4 logics.

(a) A modal public announcement logic where the identity of the actual world has several options and the public announcement narrows down these options or moves the actual world to a new set of options.

(b) A temporal logic where one can move from time t to time $t+1$ by taking an action.

(c) A combination of the two models where the need for public announcement logic arises from lack of clarity in the action at time t which is clarified by public announcement at a future time.

(d) A new type of classical Talmudic logic.

We can tell you now that in the Talmud the lack of clarity in actions comes from using the future to identify the elements to which the action is applied. So we need a Talmudic theory of individual objects.

Thus Talmudic law has inherent situations of vagueness and under-determination, emerging from a role a contingent future plays in the very definition of Halakhic states.[69]

(e) We need a model of backwards causality, as the future identification of such entitites has influence into the past, see [14].

So our logic will have four components put together as required by the Talmud.

We shall see further that the Talmud puts in some significant twists!

EXAMPLE 39 (Bookstore). John calls a bookstore and buys a book. He tells the owner I want to give one of these three translations of the Bible as a Bar-Mizva present to my nephew. I don't care which one, maybe b_1, b_2 or

[69] Within deliberations of the jurisprudential system of Halacha, Halakhic State may refer to a state of affairs in a civil, economic (*Mamonot*) or criminal law (*Nefashot*). It can also denote an evoked condition of ritual impurity(*Tumah*) or the status of a sacrificial animal (*Korban*), agronomical produce with an intermediary status between personal and consecrated property (*Trumot U'Maasrot*). All of these realms of Halacha and more have a possibility-spectrum within them, and the Halakhic state is normally the result of human action taken, or an evolved physical situation, with a predefined Halakhic meaning attached to them. In short — a Halakhic state denotes the definition, the status of an object, person or occurrence in accordance with Halacha as it pertains to personal, public, secular and religious laws (all falling under different purviews of Halakha). A popular term in the modern Yeshiva (the Brisk-Yeshiva style of learning), is Chalut, sometimes pronounced *Chalus* or *Chalos* — literally "aplication", short for "application of a law".

b_3. Here is my Visa number, you choose the book and wrap it up nicely.

Scenario 1
The bookshop puts aside copy b_1 to be posted to John. In this case Sold(b_1) holds.

Scenario 2
The bookshop does not put a copy aside. Next day someone comes to the shop personally and buys a copy. He is given b_1. Again some other guy buys another copy, he is given b_2. At this moment, copy b_3 is the copy that was sold to John. So according to this scenario, at the beginning, we have that a book was sold to John but not any specific book b_i, we cannot say that b_i was the book which was sold to John.

Now the situation can get more complicated legally. Suppose the shop is burnt down. The owner claims from the insurance the value of the three books. According to Scenario 1, the insurance company might say that book b_1 belongs to John and he should claim from his own insurance. According to Scenario 2, we may take a different view.

We can turn this problem round.

Suppose one book is damaged. Can the bookstore sell the other two books for immediate cash and tell John "Sorry, your book was damaged, please wait for a replcement"?

The story can be significantly changed if John buys shares in the shop. The shop has the only three surviving copies of the original first print of the Bible. In this case, he owns part of every book, or perhaps one of the books, without clear specification which one.

Talmudic law deals with such temporal legal complicated scenarios and to model such scenarios, we need a temporal logic with what we call the Talmudic classical models and Talmudic public announcement logic.

Summary

The situation in Figure 13 arises in the Talmud when an action is taken relating to an individual x whose identity is determined at a future time. Therefore

1. Start with a model t
2. Action **a** is performed on the x such that $\varphi(x)$ is true in the future.
3. Therefore action **a** is not completely clear. We need to wait until the identity of x is revealed in time. The possibilities are s_1, s_2, s_3.
4. This is the clarification we are waiting for. We are, of course, in difficulty in the interim period until the clarification is revealed. We

stress again that the action takes place at time t with the intention that its effect takes place immediately at t. The action needs to be clarified so we wait. When clarified, we still want the result of the action to take place at the original time t!

5. When the identity is revealed, we find ourselves in some further difficulty, to be explained later.

4.2 Talmudic classical models and Talmudic public announcement frames

We saw in the previous section that there can be a lack of clarity about the result of applying an action to a classical model. We presented schematically this lack of clarity in Figure 14.

Classical model theory can handle the fine distinctions required by Figure 14.

Our starting model is model $\mathbf{m}_0$ in Figure 14 and after action $\mathbf{a}$ we move to a triple option

$$\{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3\}.$$

Thus we can say that an extended classical model is a set of several classical models, i.e. an S5 modal model.

The Talmud looks at additional options for a model, call it a "Quantum-like" model $\mathbf{m}_4^Q$, where the predicate P is spread over the vector element (a, b, c). So in this model we have

$$(*) \qquad \neg P(a) \wedge \neg P(b) \wedge \neg P(c) \wedge P(a, b, c)^{70}$$

We say "Quantum-like" because this possibility is actually identical to the quantum superposition idea.

We therefore need to do two things:

1. increase our stock of basic models;

2. Specify how our old actions apply to the new models of (1).

Let us allow for products of models.

Consider a model

$$\mathbf{n} = \mathbf{m}_1 \times \mathbf{m}_2 \times \mathbf{m}_3.$$

[70] The predicate P is one place, the vector element (a, b, c) is a single element, so we should write

$$P((a, b, c)).$$

We abuse notation and write

$$P(a, b, c).$$

The elements of $\mathbf{n}$ are vectors (x, y, z), x from $\mathbf{m}_1, y \in \mathbf{m}_2$ and $z \in \mathbf{m}_3$. We identify the old elements a, b, c as the diagonal (pure state) vectors

$$\begin{array}{rcl}\bar{a} & = & (a, a, a) \\ \bar{b} & = & (b, b, b) \\ \bar{c} & = & (c, c, c)\end{array}$$

take the element

$$\bar{x} = (x_1, x_2, x_3)$$

Let $\|\mathbf{n}\|_i = \mathbf{m}_i$, the ith component of $\mathbf{n}$. Similalry $\|\bar{x}\|_i = x_i$. We can define

$$\mathbf{n} \vDash P(\bar{x}) \text{ iff for all } i \text{ we have } \|\mathbf{n}\}_i \vDash \|x\|_i.$$

So, for example, we have

$$\mathbf{m}_1 \times \mathbf{m}_2 \times \mathbf{m}_3 \vDash P(a, b, c)$$

but

$$\mathbf{m}_1 \times \mathbf{m}_2 \times \mathbf{m}_3 \vDash \neg P(a, a, a) \wedge \neg P(b, b, b) \wedge \neg P(c, c, c).$$

This is an "implementation" of (*) above.

Note that the product is ordered. a is chosen from $\mathbf{m}_1$, b from $\mathbf{m}_2$ and c from $\mathbf{m}_3$, i.e. each x is chosen from the model where $P(x)$ is made true.

We can now replace Figure 14 by Figure 15.[71]

We can therefore give the following definition:

DEFINITION 40 (Talmudic classical models).

1. Let $\mathbf{m}_i$ be a classical model over the same domain D for the same language $\mathbf{L}$.

2. A basic Talmudic model for $\mathbf{L}$ is any product $\mathbf{n} = \prod_{i=1}^n \mathbf{m}_i$ over the domain D^n. Define satisfaction by

$$\mathbf{n} \vDash \varphi(\bar{x}) \text{ iff for all } i\ \mathbf{m}_i \vDash \varphi(x_i), \text{ where } \bar{x} = (x_1, \ldots, x_m).$$

   An ordinary classical model $\mathbf{m}$ over domain D can be identified with any $\mathbf{m}^n$ over D^n with $a \in D$ identified with $\bar{a} = (a, \ldots, a) \in D^n$.

DEFINITION 41 (Talmudic $\mathbf{K}$ frame). A Talmudic $\mathbf{K}$ frame has the form $(S, \mathcal{R}, \mathbb{P})$ where

1. $S \neq \varnothing$ is a set of possible worlds

[71] Note that according to Remark 49 below, we shall take $\mathbf{m}_0$ here and not $\mathbf{n}_0$ as in this figure. The two are equivalent.

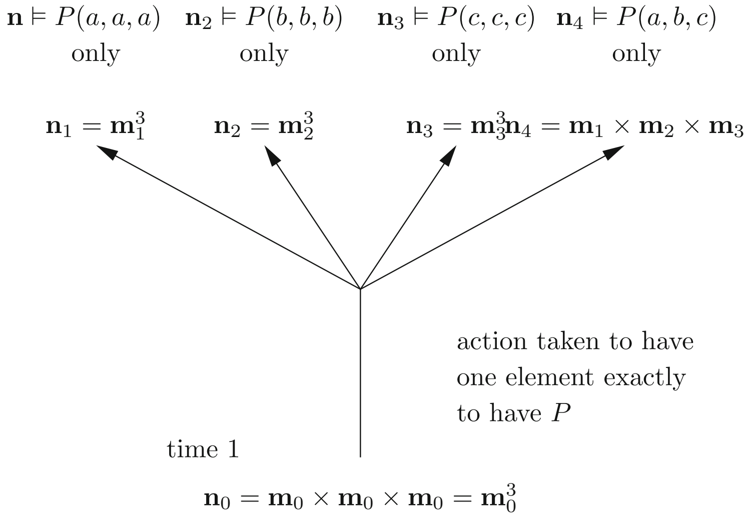


Figure 15. Talmudic classical models

2. $\mathcal{R}$ is a multi-valued accessibility relation of the form $x\mathcal{R}\{x_1, \ldots, x_n\}$, reading: one of x_i is accessible to x but we do not know which one and we await a public announcement clarification.

   Thus $\mathcal{R} \subseteq S \times S^*$ where S^* is the set of all finite subsets of S.

3. $\mathbb{P}$ is a set of public announcements of the form

$$\alpha = (x, \{x_1, \ldots, x_n\}, y), y \in \{x_1, \ldots, x_n\}$$

   where

$$(x, \{x_1, \ldots, x_n\}) \in \mathcal{R}$$

   reading:
   I hereby announce that y is the element accessible to x.

   Given α, let $\|\alpha\|$ be $(x, \{x_1, \ldots, x_n\})$.

   The above is a deterministic public announcement, because it chooses exactly one $y \in \{x_1, \ldots, x_n\}$. The public announcement can be non-deterministic if it chooses a subset Y of $\{x_1, \ldots, x_n\}$. It therefore has the form $\alpha = (x, \{x_1, \ldots, x_n\}, Y)$, Y is a subset of $\{x_1, \ldots, x_n\}$. We allow Y to be empty.

4. For a given α and $\mathcal{R}$, let $\mathcal{R}_\alpha$ be

$$\mathcal{R}_\alpha = (\mathcal{R} - \|\alpha\|) \cup \{(x, Y)\}$$

   where $\alpha = (x, \{x_1, \ldots, x_n\}, Y)$.

5. Let $\bar{t} = (t_1, \dots, t_n)$ be a sequence of points in S. We define by induction the notion of $\bar{t}$ being a legitimate sequence from t_1 to t_n.

   5.1. (t_1, t_2) is a legitimate sequence if for some $T_2 \subseteq S$ we have $(t_1, T_2) \in \mathcal{R}$ and $t_2 \in T_2$.

   5.2. $(t_1, \dots, t_{n+1})$ is a legitimate sequence if (t_1, t_2) and $(t_2, \dots, t_{n+1})$ are legitimate.

6. A legitimate sequence $(t_1, \dots, t_m)$ is said to be *clarified* if $(t_i, \{t_{i+1}\}) \in \mathcal{R}$, for $i = 1, \dots, n-1$.

7. Let $\alpha_1, \dots, \alpha_k$ be public anouncment from $\mathbb{P}$. Let $\bar{t} = (t_1, \dots, t_n)$ be a legitimate sequence from t_1 to t_n. We say that $\alpha_1, \dots, \alpha_k$ clarify $\bar{t}$ if $\bar{t}$ is clarified in $\mathcal{R}_{\alpha_1, \dots, \alpha_n}$.

DEFINITION 42 (Talmudic **K** models and Talmudic **K** syntax). We can derive Talmudic **K** models from Talmudic **K** frames of Definition 41 as follows:

1. Let the set of possible worlds S^* be the set of all legitimate sequences of the frame.

2. Let the accessibility relation R^*, which is dependent on $\mathcal{R}$, be defined by $\bar{t} R^*_{\mathcal{R}} \bar{t}'$ iff for some s in S we have $\bar{t} = (t_1, \dots, t_n)$ and $\bar{t} = (t_1, \dots, t_n, s)$.
   Let h be an assignment to the atoms.

3. We define the syntax of the language. We have atoms, the classical connectives and modal operators as follows:

   - a necessity operator $[N]$.
   - for $\alpha = (x, \{x_1, \dots, x_n\}, Y)$ being a public announcement, define the operator $[\alpha]$.

   Let $[N]$ be a necessity operator and A a formula.

   Define
   $[N]A$ holds at x in (S^*, R^*) iff for all y such that $x R^*_{\mathcal{R}} y$, we have that A holds at y.

4. Let $\alpha = (x, \{x_1, \dots, x_n\}, Y)$ be a public announcement, define
   $[\alpha]A$ holds in (S^*, R^*) at x iff [(If x is a legitimate sequence in (S, R_α) then A holds at x in $(S^*, R^*_{(\mathcal{R}_\alpha)})$.

5. Note that the syntax depends on the particular semantics and contains elements from the semantics.

REMARK 43. Note that Definition 42 is not adequate for the purpose of modelling Talmudic behaviour. The problem is with clause 4. Suppose we have a legitimate sequence $\bar{t} = (t_1, \ldots, t_n)$ and assume that t_2 is taken from the set $T_2 = \{t_2, s\}$ such that $(t_1, T_2) \in \mathcal{R}$. Further suppose that it is clarified that the correct new state after t_1 should be s. Clause 4 does not apply to this public announcement because *$\bar{t}$ is no longer a legitimate sequence. So we do not have a semantics for this case.*

We must provide a new legitimate sequence to replace $\bar{t}$.

So what is this new legitimate sequence?

The answer is not clear because we cannot say the obvious, namely that it is $\bar{t}' = (t_1, s, t_3, \ldots, t_n)$, because this $\bar{t}'$ may not be a legitimate sequence.

Since a legitimate sequence indicates a possible world where we are at a certain time, then if the public announcement clarification literally cancels that sequence, we need to know where we are going to be after the announcement!

This question still needs still to be addressed. We shall do this in Section 4 leading to Definition 53 and we shall further give full discussion and comparison in Section 5

REMARK 44. We quickly compare our Talmudic public announcement models frames with the traditional one. See Appenidx B and [40, Chapter 4]. A more detailed comparison and discussion is done in Section 5.

1. Traditional public announcement logic operates as follows. We have a modal **K** model (S, R, t) and we are at node t. We announce a wff φ such that $t \vDash \varphi$. We move to the new model $(S_\varphi = \{s | s \vDash \varphi\}, R, t)$.[72]

2. Talmudic public announcment logic we have $(S, \mathcal{R}, t)$. We announce α and we move to $(S, \mathcal{R}_\alpha, t)$.

3. In traditional (say constant domains) public announcement logics the models $\mathbf{m}_t$ associated with t are classical models.

4. In Talmudic public announcement logic the models $\mathbf{n}_t$ associated with t are products, as in Definition 40.

   Furthermore, the assignment is not arbitrary but respects the geometry of $(S, \mathcal{R})$ the details of what this means to be defined later.

DEFINITION 45.

1. Let $(S, \mathcal{R}, \mathbb{P})$ be a deterministic Talmudic **K** frame. A subset $\mathbb{P}_0 \subseteq \mathbb{P}$ is said to be consistent if for no $\alpha, \beta \in \mathbb{P}_0$ do we have

$$\alpha = \{x, \{x_1, \ldots, x_n\}, y) \text{ and}$$
$$\beta = (x, \{x_1, \ldots, x_n\}, z\} \text{ and } y \neq z.$$

[72]If φ is announced and it is not true at t then we do nothing.

2. Let $\mathbb{P}_0$ be consistent, then define $\mathcal{R}_{\mathbb{P}_0}$ to be

$$\mathcal{R}_{\mathbb{P}_0} = \mathcal{R} - \{\|\alpha\| | \alpha \in \mathbb{P}_0\} \cup \{(x,y) | (x, \{x_1, \ldots, x_n\}, y) \in \mathbb{P}_0\}.$$

   Note that in the above we abuse notation and identify (x, y) with $(x, \{y\})$.

3. $\mathbb{P}$ is said to be a properly clarifying set iff for every maximal consistent subset $\mathbb{P}' \subseteq \mathbb{P}$ we have that $\mathcal{R}_{\mathbb{P}'}$ is a binary relation

$$\mathcal{R}_{\mathbb{P}'} \subseteq S \times S.$$

4.3 Introducing Talmudic temporal public announcement logic

We now introduce time into our models. Our concept of time is discrete and time ticks discretely as we move from one state to another by executing some action. So Figure 16 is a classic discrete flow of time

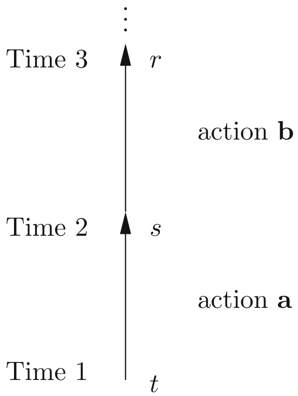


Figure 16. Time and actions sequence

The problem arises when the actions involve the need for clarification. Then we get Figure 17.

Note that in Figure 17 we move from t to s by executing action $\mathbf{a}$. Now because of lack of clarity about action $\mathbf{a}$, we might end up at states $s_1, \ldots, s_n$. We now apply action $\mathbf{b}$. We therefore must apply $\mathbf{b}$ to each of the states $s_1, \ldots, s_n$. We thus get the possible states $r^i_1, \ldots, r^i_m$. Note that the same action $\mathbf{b}$ is applied to each option $s_1, \ldots, s_n$ and that the resulting split is to m options for each s_i (i.e. m depends on $\mathbf{b}$ only and not on i). This is a design assumption motivated by the Talmud and not by any technical reason!

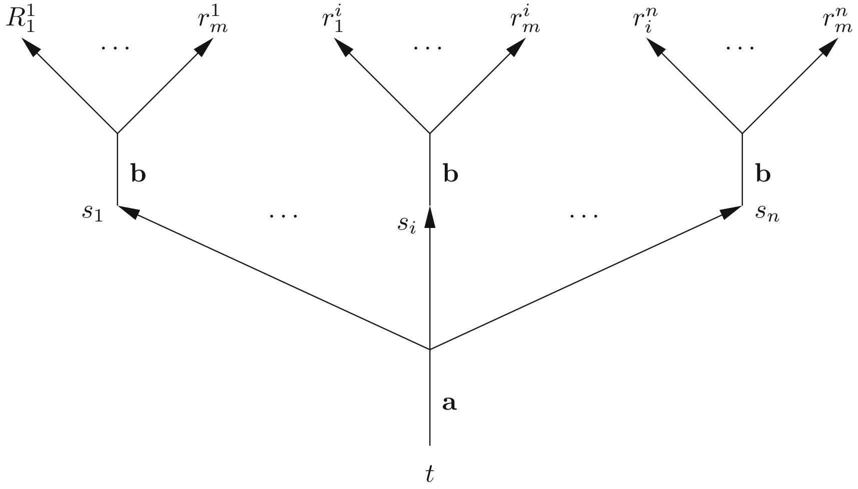


Figure 17. Lack of clarity in time and actions sequence

Also note that our system must tell us how to apply **b** to cases where the models at state s_i are new superposition models.

EXAMPLE 46 (How to apply actions to superposition models). Let us go and take another look at Figure 15. Suppose we have another predicate $\lambda x P_1(x)$, which we want to apply to one of the elements in $\{a, b, c\}$. Suppose in the initial model $\mathbf{m}_0$ we have

$$\neg P_1(a) \wedge \neg P_1(b) \wedge \neg P_1(c).$$

We can easily apply first say $P(a)$ and then go on and apply say $P_1(b)$. The result is the model with

$$\neg P_1(a) \wedge P_1(b) \wedge \neg P_1(c) \wedge P(a) \wedge \neg P(b) \wedge \neg P(c).$$

Now if the P-application is not clear, then we get the four options of Figure 15. We now want to apply the $P_1(b)$ action. Since we have four optional models $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ and $\mathbf{n}_4$, we will have to say that we apply $P_1(b)$ to each one of them. What we do in the case of $\mathbf{n}_1$–$\mathbf{n}_3$ is clear. We apply $P_1(b, b, b)$ to each. But what do we do in the case of $\mathbf{n}_4$? Well, you may think what is the problem? In $\mathbf{n}_4$ let us apply **b** and have

$$P(a, b, c) \wedge P_1(b, b, b).$$

This formulation might be sufficient, but now suppose that we have the integrity constraint

(IC) $\neg\exists x[P(x) \wedge P_1(x)]$.

Now in this case we cannot apply $P_1(b,b,b)$ to $\mathbf{n}_2$. But can we apply it to $\mathbf{n}_4$?

Does $P(a,b,c) \wedge P_1(b,b,b)$ violate the integrity constraint?

We can say no, (a,b,c) and (b,b,b) are not the same x or we can say that the property P is superimposed also on b in (a,b,c) and therefore we adopt the view that we cannot apply $P_1(b)$.

The Talmud will choose the option depending on the meaning of P and P_1.

We can say do not apply $P_1(b)$. We can also say apply $P_1(b)$ and retract if the clarificiation for P chooses $P(b)$.

We shall see how the Talmud deals with these cases in a later section.

Our formalism must be able to deal with these options.

DEFINITION 47. Let $(S, \mathcal{R}, \mathbb{P}, t)$ be a model

1. The model is said to be discrete linear if for any $s \in S$ there exists a unique $T \subseteq S$ such that $(s,T) \in \mathcal{R}$.
2. t is said to be a root point (starting point) if for every $s \in S$, there exists a unique legitimate chain from t to s.
3. The model is potentially linear if for every maximal consistent $\mathbb{P}' \subseteq \mathbb{P}$, $\mathcal{R}_{\mathbb{P}'}$ is a discrete linear chain with t as its first element.

DEFINITION 48 (Embedding of Talmudic models). Let $\mathbf{n}$ be a classical Talmudic model of the form $\mathbf{n} = \prod_{i=1}^{r} \mathbf{m}_i$, where each $\mathbf{m}_i$ is an ordinary classical model for the same language based on the same domain D.

Let $\mathbf{n}_1, \ldots, \mathbf{n}_k$ be k such models, i.e. $\mathbf{n}_j = \prod_{i=1}^{r} \mathbf{m}_i^j$, where each $\mathbf{m}_i^j$ is a classical model based on the same domain D. Note the "r" is fixed for all models. We define the diagonal embedding of $\mathbf{n}$ into $\mathbf{n}^* = \prod_{j=1}^{k} \mathbf{n}_j$.

1. Each element $\bar{a}$ in the domain of $\mathbf{n}$ is mapped onto $(\bar{a}, \ldots, \bar{a})$ in $\mathbf{n}^*$.

   We also have $\mathbf{n}^* \vDash \varphi(\bar{x})$ iff for each $j, \mathbf{n}_j \vDash \varphi(\|\bar{x}_j\|)$.

REMARK 49. Consider again Figure 15. In this figure $\mathbf{n}_0 = \mathbf{m}_0^3$ is embedded into $\mathbf{n}_1, \ldots, \mathbf{n}_4$. If we were to conform to what is suggested in Definition 48, we would take $\mathbf{m}_0$ in the figure and not $\mathbf{m}_0^3$. This simplifies the presentation and complexity of our models. Note that $\mathbf{m}_0$ and $\mathbf{m}_0^3$ are basically the same model since satisfaction is achieved coordinatewise!

DEFINITION 50 (Connection between models). Assume $(x, \{x_1, \ldots, x_n\}) \in \mathcal{R}$ and assume $\mathbf{m}_x^0$ is a Talmudic classical model associated with x. Let D_x be its domain. We now say what models we associate with $x_1, \ldots, x_n$. We need to assume an action $\mathbf{a}$ which when applied to $\mathbf{m}_x$ yields k possible outcomes

$$\mathbf{m}_x^1, \ldots, \mathbf{m}_x^k.$$

Let $\mathbf{n}$ be any product

$$\mathbf{n} = \prod_{j=1}^{k} \mathbf{n}_i$$

where $\mathbf{n}_i \in \{\mathbf{m}_x^1, \ldots, \mathbf{m}_x^k\}$ and let $z \in \{x_1, \ldots, x_n\}$.

Then we can take the model $\mathbf{n}$ to be the model at z, i.e. $\mathbf{m}_z = \text{def}.\mathbf{n}$.

Clearly we can embed $\mathbf{m}_x$ into $\mathbf{m}_z$ as in Definition 48.

EXAMPLE 51 (Talmudic views). We are now ready to list the Talmudic opinions about the situation in Figures 14 and 15. Think of the action as legally endowing entity x with the legal status $\lambda x P(x)$. The action was not clear and the available options for clarification are $\mathbf{n}_1, \ldots, \mathbf{n}_4$.

The simplest story we can give is that we want to confer status P now at time 1 on the x such that $x \in \{a, b, c\}$ and x wins the race to take place at time 7.

Obviously we need to wait for time 7 to clarify the situation. Meanwhile, the following can happen:

1. Nothing happens at time 7.
2. a wins at time 7.
3. b wins at time 7.
4. c wins at time 7.
5. a dies at time 2.
6. a and b die before time 5.

Obviously from the logical point of view we get lack of clarity[73] because we define at time $t = 1$ an entity using the Iota $ix\varphi(x)$, where φ is a predicate dependent on time 7.

Formally we have

$$(\sharp) \qquad [y \in \{a, b, c\} \wedge [y = ix\varphi(7, x)] \to P(1, y)$$

The question is: Do we accept such definitions?

These are called *Breira* (*choice*)[74] in the Talmud.

We have the following Talmudic positions

[73]This problematic feature of future indeterminates has been recognised in classical aristotelian logic and was treated along the lines of binary truth values that apply to all times, [1]. In pp. 140-142 the different explanations for the status of the predicate of a future (and later past) event in Aristotelian logic.

[74]Breira (literaly 'determination'/'resolution') is an underdetermined or uniquely future-oriented choice defining a Halakhic state, that differs from a regular condition (Tenai). Breira is chiefly discussed in Trachtate Gittin 25a, 74a concerning divorce law, and Trachtate Eruvin 37b regarding definitions of extended 'personal space' in holidays (allowing for motion beyond the default degrees of freedom). A more involved case appears in Trachtate beitza 10a, in the context of deciding on a specific fowl for the holiday feast. For a treatment of classic logical attributes of Breira, [78].

1. We do not accept such definitions. Nothing happens. Reject i.e., making such a conditional proposition is not logically coherent and carries no sense.

2. We do not accept such definitions but nevertheless something does happen. We move to the quantum model $\mathbf{n}_4$ *immediately* from time 2, and even at time 7 if a specific y is found, we still remain with model $\mathbf{n}_4$.

   Of course, if say c dies at time 2 then from time 3 the superposition may be on $\{a, b\}$ only, but not necessarily. For example if it is not clarified to which of $\{a, b, c\}$ I am married, then if c dies then I may be considered a widower, so the superposition (a, b, c) still continues. In the case of the book shop with the three books, it is possible that one book is burnt and destroyed and nevertheless it is the one which was sold. So legally it is still there for insurance claims etc. In fact, in legal and every day life elements never die, we still talk about them. In the UK it is possible for parents to register their unborn child to Beavers (the youngest age group in the Scouting movement) on the expectation of the child's coming existence.

   To be quite clear, this position says that even when there is clarity at time 7 that $P(a)$ should hold, we still stick with the (quantum-like) superposition $P(a, b, c)$. So if at time 8 we want to execute $P_1(b)$, with the constraints

   $$\neg\exists x(P(x) \wedge P_1(x))$$

   we still reject the action because the superposition on b remains. See Example 46.

3. We do accept such definitions. We wait for time 7 for clarification to find the y (if it exists) and $P(y)$ holds from time 1.

4. We accept the definition, however the clarification is effective only from time 7. At the intermediate times, times 2–6, the model is the (quantum-like) superposition model $\mathbf{n}_4$. Again if c dies the superposition is reduced. Even after time 7 when we look back and ask what model was at time 2? We will say $\mathbf{n}_4$.

5. We accept the definition and the clarification is backward causal, i.e. once clarified it investigates a Halakhic state that starts from time 1.

What is the difference between (3) and (5)?

Suppose at time 6 we cancel the practice of conferring $\lambda x P(x)$ on people. According to (3), y already got $P(y)$ except that we had to wait for time 7 to clarify who y is.

According to (5), the action takes place at time 7, when we know who y is, and the action is conferred backwards in time. Since at time 6 we cancelled this practice, $P(y)$ will not happen!

Note that using the terminology and model in [14], (3) is the Rabbi Fisher approach, while (5) is the Rabbi Shkop approach. See Example 13 of [14].

REMARK 52 (Methodological comments). The perceptive reader must have noticed that so far we defined semantic models to help us understand Talmudic reasoning. We gave Kripke type semantics but did not give any corresponding formal syntax. This is not needed in principle. Think for example of Situation semantics, pioneered by Jon Barwise and John Perry in the early 1980s, this was a purely semantic attempt to provide a solid theoretical foundation for reasoning about common-sense and real world situations. No axiom system was necessary. Only a sharpening of concepts using formal semantics of logic.

So to explain the various opinions and nuances of the great Rabbis about lack of clarity in action and time the semantics is sufficient, and the various distinctions can be written in English in the metalevel. No need to introduce modal operators and further write the distinctions in the language of the operators and prove a completeness theorem.

Nevertheless, this new Talmudic semantic does inspire new types of logics and we shall present some in Section 4.

4.4 Propositional Talmudic public announcement logic **TPK**

We now introduce propositional public announcement logic based on **K** inspired (see Remark 52) by the Talmud. We first need to motivate the formal design of the system.

4.5 Motivation

Consider a state t at time 1. To have a concrete example, assume John owns a certain book. John performs an action $\mathbf{a}$ depending on the future. He gives the book either to Tracy or to Mary, provided that next week, at time 7, a coin is fipped. If it lands heads then the book is Mary's (action $\mathbf{a}_1$ at state s_1) and if lands tails, then it is Tracy's (action $\mathbf{a}_2$ at state s_2).

Now consider another action:

$$\mathbf{b} = \text{Tracy writes her name in the book.}$$

Its precondition is that $\alpha =$ Tracy owns the book.

Its postcondition is

$$\beta = \text{Tracy's name is in the book.}$$

The question is: Can Tracy perform action **b** at time 2?

Well, if at time 7 Tracy wins, then the action **b** at time 2 is OK, but if not then the precondition of the action is not fulfilled and so **b** cannot be performed. However, at time 2 we do not know who owns the book. So one of two scenarioes can be allowed to happen at time 2:

1. **b** is not allowed to be executed.
2. **b** is tolerated, i.e. it can be executed anyway and a risk is taken.[75]

Note that no matter what the policy is, it is a symmetrical policy, as Figure 18 shows, with respect to states $\{s_1, s_2\}$.

1. Either **b** cannot be executed, neither at s_1 nor at s_2; or
2. **b** is tolerated both at s_1 and at s_2.

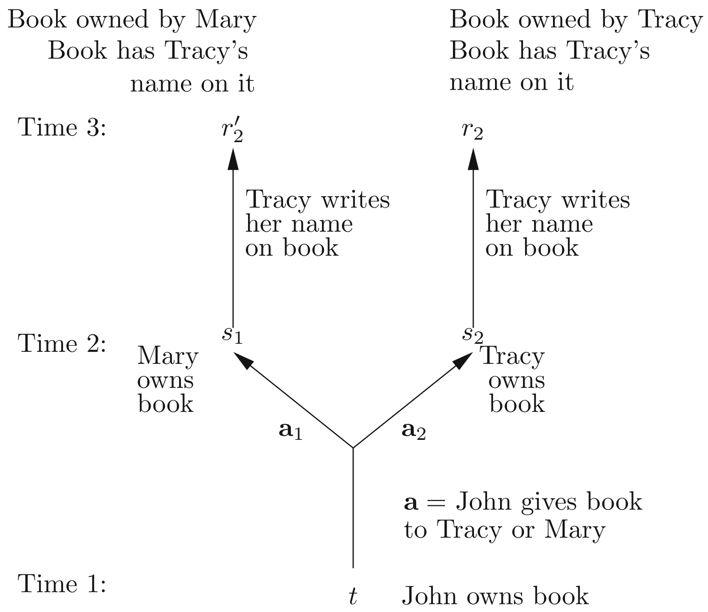


Figure 18. Actions which are tolerated

The important feature of Figure 18 is that action

[75] The Talmud deals with situations like this, where there is a doubt. The Bible requires us to be strict and so we should not tolerate Tracy writing her name on the book. The situation of flipping a coin is not under her control. Suppose for comparison that the book is given to Tracy on condition that at time 7 she cleans her flat, and if she does not do so then the book goes to Mary. Tracy can argue at time 2 that she is in control and at time 7 she will indeed clean her flat. So her action at time 2 of writing her name in the book, may be tolerated.

$$\mathbf{b} = \text{Tracy writes her name on book}$$

is tolerated even if its precondition does not hold. This is because the states $\{s_1, s_2\}$ are regarded as some sort of superposition single entangled state $s_1 \times s_2$. So either $\mathbf{b}$ can be applied to all of them or to none of them.

The technical importance of this observation can be seen in Figure 19.

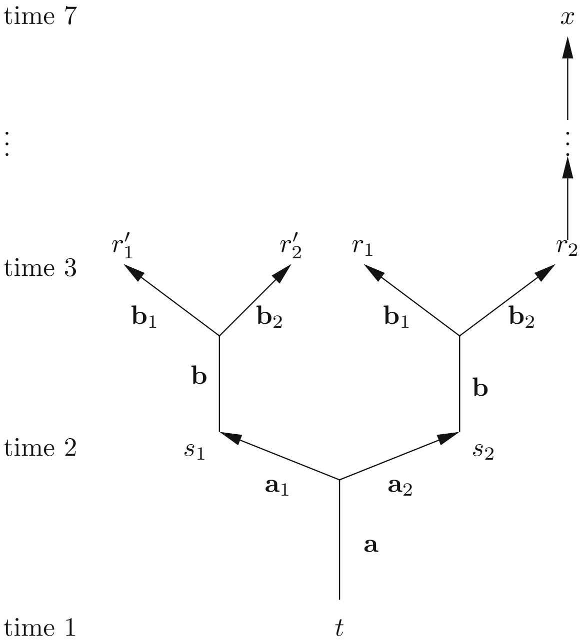


Figure 19. Technical view of tolerated actions

In Figure 19 the two actions $\mathbf{a}$ and $\mathbf{b}$ are not clear and require clarification. $\mathbf{a}$ can be $\mathbf{a}_1$ or $\mathbf{a}_2$ and $\mathbf{b}$ can be $\mathbf{b}_1$ or $\mathbf{b}_2$. Suppose Tracy uses disappearing ink which holds for a maximum of 3 days. So there are two possibilities:

1. Tracy's name is permanently on the book.
2. Tracy's name is not permanently on the book.

Suppose at time 7 we discover that Mary is the owner. This is the public announcement. Suppose that at time 7 we are at node x, because we chose the path $(t, \mathbf{a}_2, s_2, \mathbf{b}_2, r_2, \ldots, x)$ at our own best guess and risk.

The public announcement says we should have gone to $(t, \mathbf{a}_1, s_1, \ldots)$. So where are we now? We are not at x. If we allow for $\mathbf{b}$ to be executed at s_1 as well as at s_2, with the same possible options, then we can continue the same continuation path from s_1.

To do this we must pair the unclear states r_1, r_2 which are the result of the lack of clarity when $\mathbf{b}$ is executed on s_1 with r_1' and r_2', which are the result of $\mathbf{b}$ executed at s_2.

We can thus continue

$$(t, \mathbf{a}_2, s_1, \mathbf{b}_2, r_2', \ldots, y)^{76}$$

So what are the formal assumptions we need on our formal modelling?

1. Any action $\mathbf{a}$ has the lack of clarity that it might be $\mathbf{a}_1, \ldots, \mathbf{a}_k$. When in state t we perform action $\mathbf{a}$, then we might be in any of the states $s_1(t, \mathbf{a}_1), \ldots, s_k(t, \mathbf{a}_k)$.

2. Any other unclear action $\mathbf{b}$, which might be $\mathbf{b}_1, \ldots, \mathbf{b}_m$ can be applied to any of $s_i(t, \mathbf{a}_i)$ resulting in $r_j(s_i(t, \mathbf{a}_i), \mathbf{b}_j)$ $1 \leq j \leq n$.

   The number of outcomes m is fixed and depends on $\mathbf{b}$ only, and not on $s_i(t, \mathbf{a}_i)$.

3. The m outcomes for each $s_i(t, \mathbf{a}_i)$ are matched and the listing indicates the matching. Thus for each $1 \leq j \leq n$ $r_j(s_1(t, \mathbf{a}_1), r_j(s_2(t, \mathbf{a}_2), \ldots,$ $r_j(s_k(t, \mathbf{a}_k))$ are matched because they are all obtained by the application of action $\mathbf{b}_j$.

Note that two matched states need not be the same. In Figure 18, r_1 is matched to r_2', but in r_2 Tracy owns the book and in r_2', Mary owns the book. The states are matched because both are the result of the action $\mathbf{b}$ of Tracy writing her name on the book in permanent ink.

The notation we use is

$$(t, \mathbf{a}_i, s_i, \mathbf{b}_j, r_j, \ldots)$$

4.6 Preliminary formal discussion

We now semi-formally motivate and explain our system. Let $\mathbf{A}$ be a set of actions. Let s_0 be an initial state. If the actions are all deterministic then we can move from state to state by applying the actions. The following is a simple run:

$$s_0 \mathbf{a}\mathbf{a}'\mathbf{a}'' \ldots$$

[76] Note that we can use simpler notation. The points in Figure 19 can be identified by the actions leading to them. Thus we can write

$$\begin{aligned} s_1 &= t\mathbf{a}_1 \\ s_2 &= t\mathbf{a}_2 \\ r_1 &= t\mathbf{a}_2\mathbf{b}_1 \\ r_2 &= t\mathbf{a}_2\mathbf{b}_2 \\ r_1' &= t\mathbf{a}_1\mathbf{b}_1 \\ r_2' &= t\mathbf{a}_1\mathbf{b}_2. \end{aligned}$$

We shall use this notation in Section 4.2.

The state arising from $_0$ after the application of action $\mathbf{a}$ is $s_1 = (s_0\mathbf{a})$, etc, etc.

Actually what we have here is a deterministic automaton, where the states are sequences of actions applied to s_0 and the alphabet are the actions.

Now assume that the actions $\mathbf{a} \in \mathbf{A}$ have some ambiguity to be clarified at a later stage (i.e. after more actions are applied). So applying $\mathbf{a}$ could be any of $\mathbf{a}_1, \ldots, \mathbf{a}_{\mathbf{k(a)}}$, where $\lambda\mathbf{x}\mathbf{k}(\mathbf{x})$ gives us the number of possibilities for $\mathbf{x}$.

Then when we apply say $s_0\mathbf{a}\mathbf{a}'\mathbf{a}''$ we can get any one of the states

$$\{s_0\mathbf{a}_i\mathbf{a}'_j\mathbf{a}''_r\}$$

where $1 \leq i \leq k(\mathbf{a}), 1 \leq j \leq k(\mathbf{a}')$ and $1 \leq r \leq k(\mathbf{a}'')$.

If at this point there is a public announcement that $\mathbf{a}$ is indeed $\mathbf{a}_1$, then part of the ambiguity is resolved and our options are now

$$\{s_0\mathbf{a}_1\mathbf{a}'_j\mathbf{a}'_r\}$$

We are now ready to define the model. We do this in stages.

1. A model has the form $(S, \mathbf{A}, s_0, \rho, h)$ where $\mathbf{A}$ is a set of actions and S is a set of elements of the form

$$\alpha = (s_0\mathbf{a}^1_{j_1}\mathbf{a}^2_{j_2}\ldots\mathbf{a}^m_{j_m})$$

   where $1 \leq j_i \leq k(\mathbf{a}^i), \mathbf{a}^i \in \mathbf{A}$ and $1 \leq \mathbf{k}(\mathbf{a}^i)$.

   We also allow for $\alpha = (s_0)$.

2. Let R_1 be the relation $\alpha R_1 \beta$ iff $\beta = \alpha * (\mathbf{a}_j)$ for some $a \in \mathbf{A}$ and $j \leq \mathbf{k}(\mathbf{a})$. Let R be the transitive closure of R_1.

3. We need the notion of a node x is at a distance n from (s_0).

   - (s_0) is at a distance 0 from itself.
   - If α is at a distance n from (s_0) then $\alpha * (\mathbf{a})$ is at a distance $n+1$.
   - The distance is actually the time, since we are applying the actions in sequence, one after the other.

4. We now define ρ. Let $\alpha = \beta * (\mathbf{b}_j) * \gamma$ where we have that $\mathbf{b}_j$ is one option from among $\{\mathbf{b}_1, \ldots, \mathbf{b}_{\mathbf{k(b)}}\}$.

   We are at point $\alpha = \beta * (\mathbf{b}_j) * \gamma$. At this point there is a public announcement that the correct meaning of action $\mathbf{b}$ taken at the β level was $\mathbf{b}_1$ and not $\mathbf{b}_j$. So we actually should have been at $\beta * (\mathbf{b}_1)$ and subsequent actions γ would bring us to $\beta * (\mathbf{b}_1) * \gamma$.

   Now suppose we take action $\mathbf{e}$ and apply to our current state. Without the public announcement we move to $\beta * (\mathbf{b}_j) * \gamma * (\mathbf{e})$.

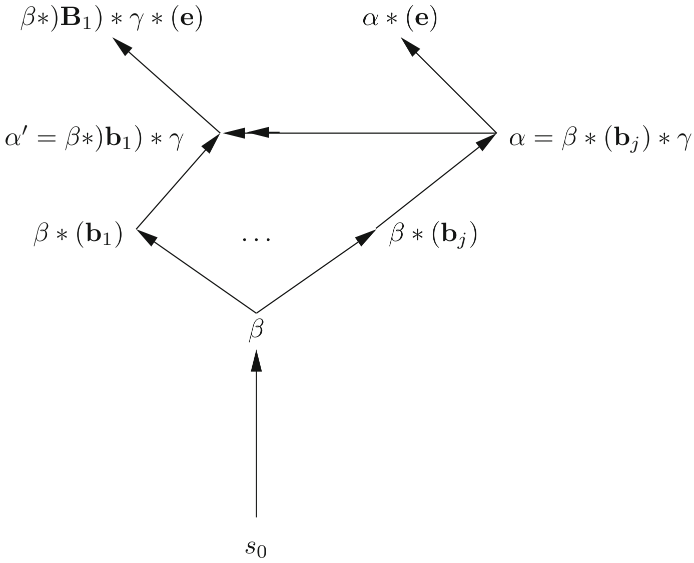


Figure 20. The function ρ

However, the public announcement says we should be at $\alpha' = \beta * (\mathbf{b}_1) * \gamma$ and so we should move to $\alpha' * (\mathbf{e}) = \beta * (\mathbf{b}_1) * \gamma * (\mathbf{e})$.

Thus the effect of the public announcement at α is to send us to $\alpha' * (\mathbf{e})$ which is $\beta * (\mathbf{b}_1) * \gamma * (\mathbf{e})$.

We draw the function ρ as in Figure 20. Ordinary arrows $\rightarrow$ indicate R and double arrows $\twoheadrightarrow$ indicate ρ.

The relation ρ between α and α', identifies the public announcement uniquely. Note that α and α' are at the same distance from (s_0).

Actually there is a problem. What if the public announcement says that actually $\mathbf{b}_j$ was the correct choice? How do we express that? We will have to let ρ take us to $\beta * (\mathbf{b}_j)$. The perceptive reader might ask, why not be consistent and let ρ take us to $\beta * (\mathbf{b}_1)$ in all cases and the case where $\mathbf{b}_1 = \mathbf{b}_j$ will sort itself automatically?

The problem with that is that we need to identify where we are going to be when we apply the next action $\mathbf{e}$, in our example we need to identify $\beta * (\mathbf{b}_1) * \gamma * (\mathbf{e})$. We are using ρ to help us so we let ρ point to $\beta * (\mathbf{b}_1) * \gamma$.[77]

[77] The pure minded reader may object to our approach. He will say that we can still identify where to go. Let ρ take us to the point $\beta * (\mathbf{b}_1)$, which is the conceptually correct clarification point. We now know two points

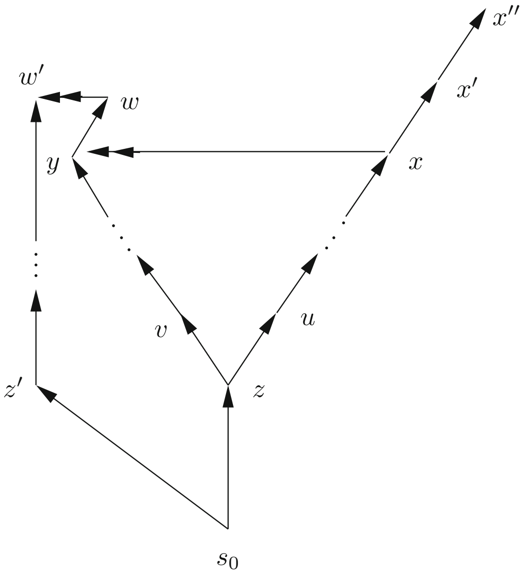


Figure 21. Public announcement and ρ

The very next paragraph does public announcement without action symbols, using just accessibility relation, and and the problem of identifying where we should be is crucial.

Let us now look at this differently. Assume $x\rho y$. Since (S, R) is a tree, let z be the unique maximal first point below both x and y. Note that if y itself is below x then z is the predecessor of y. Let u be the next point in the direction of x and v the next point in the direction of y. Again note that if y is below x then $u = v = y$. Then the public announcement at x says that u is clarified and really should be v. Again, if y is below x then the public announcement says that u was the correct choice.

We can write the public announcement in the form $x\rho y$ and we identify the point z as its *target base*.

Figure 21 describes the situation.

Note that whenever $x\rho y$ holds and x is at distance n from the root s_0, then y is at distance $n + 1$, unless y is below x.

(a) where we are, namely $\alpha = \beta * (\mathbf{b}_j) * \gamma$

(b) the clarified action $\alpha = \beta * (\mathbf{b}_1)$.

From the above two items we can identify γ and go to $\alpha' = \beta * (\mathbf{b}_1) * \gamma$.

This is true but only because points are identified by sequences of actions. We do not have this luxury in the general case.

Note that Figure 21 clarifies a further point. In this figure there is another public announcement namely $w\rho w'$. This one says that z should have been z'. Note that the public announcements need not come in the same order as the order of the ambiguities. (z, z') came before (u, v) but was clarified after. This can be a bit problematic because once we go through z' we do not pass through z any more. So we had better require that the clarifications come in order. In the notation where points are described by actions as in Figure 20, the order is not important and is not confusing because it is the actions that are clarified.

Note also that if we have another public annuncement sending say w' to an extension of x' of x (i.e. $w'\rho x'$) then a previous public announcement will be reversed.

We can require coherence and stability and not allow such reversals!

5. We have one more point to discuss. Imagine we have had a clarification as in Figure 21, where we had $x\rho y$. Take the branch of history from z to x. This is not the real history because we are moving to point y.

   However this history is real past for anyone living along the path from z to x.

   To make it real to the reader assume the action taken at z was the marriage of John and Mary using a priest who has been ambiguously ordained. John and Mary continued as a married couple from z up to x when it was announced that the priest was not properly ordained and therefore the marriage is null. So John and Mary move to point y and their history (path z to y) does not include marriage. We must allow them to remember as part of their past also the path z to x.

   Thus when we create a temporal model out of R and ρ, we must take the above into account!

4.7 *Formal* **TPK**

In view of the discussion in the previous two subsections, we are now ready to present our Talmudic public announcement logic.

DEFINITION 53 (Deterministic **TPK** model).

1. A deterministic **TPK** model has the form $(S, R_1, R, \rho, s_0, h)$ where (S, R_1, s_0) is a tree with root s_0 and successor relation R_1 and where R is the transitive closure of R_1.

2. ρ is a functional relation satisfying the following properties

2.1. $\forall xyz(x\rho y \wedge x\rho z \rightarrow y = z)$[78]

2.2. If $x\rho y \rightarrow y \neq s_0$.

2.3. $x\rho y \wedge \neg yRx \wedge y \neq x \rightarrow D(y) = D(x)$, where D is the distance from the root, $D(s_0) = 0$ and whenever uR_1v then $D(v) = D(u) + 1$.

2.4. Let $x \in S$ and let zRx. We say the successors of z are publicly clarified to be v, where zR_1v holds, if we have $x\rho y$, with vRy.

We require the property of coherence:

- If $\{zR_1v_1$ and zR_1v_2 and $[z$ is publicly clarified at x_1 to be $v_1]$ and $[z$ is publicly clarified at x_2 to be $v_2]\}$ then $v_1 = v_2$.

2.5. Let z be any point, then there exists an x, such that zRx and the successors of z are publicly clarified at x.

2.6. Since every z has a unique successor which is publicly clarified at some point, this means that there exists a path $\pi = (s_0, s_1, \ldots)$ such that for every $0 \leq i$, s_{i+1} is the uniquely clarified successor of s_i. Note that this is why we call the model deterministic.

2.7. Note that if we define, for x in S, the set T_x to be $\{y|xR_1y\}$ and let $\mathcal{R}$ be defined as the set of all pairs (x, T_x), we get a frame in the sense of Definition 41. We can also define the public announcements correctly from ρ using item 2.4.

3. We now define a temporal relation $<$ on the model. We use the notion of legitimate sequence of worlds

3.1. (s_0) is a legitimate sequence.

3.2. Assume $(s_0, \ldots, s_n)$ is a legitimate sequence leading to s_n.

Case 1. $\neg\exists x[s_n\rho x]$.

In this case let w be any point such that s_nR_1w, then $(s_0, \ldots, s_n, w)$ is a legitimate sequence.

Case 2. For some y we have $(s_n\rho y \wedge \neg yRs_n)$.

In this case let w be any point such that yR_1w, then $(s_1, \ldots, s_n, w)$ is a legitimate sequence.

Case 3. For some y $s_n\rho y \wedge yRs_n$.

Then this case is like Case 1.

3.3. Define $x < y$ iff there exists a legitimate sequence $(s_0, \ldots, s_i, x, s_{i+2}, \ldots, y)$.

REMARK 54 (Axiomatic formulation of **TPK**). It may be of interest to the reader to see how we can design syntax and axioms for the deterministic

[78] 1. There may be no z such that $x\rho z$.
2. This condition is for deterministic actions only, otherwise ρ is just a binary relation.

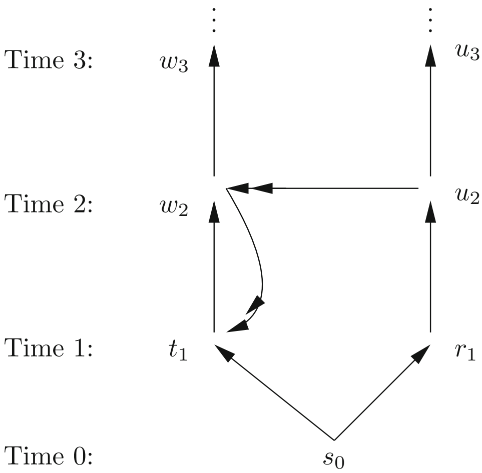


Figure 22.

semantics of Definition 53. The models involve the tree relation R_1 and the functional relation ρ. Each one of these would require a modality to control it. To motivate the syntax, let us give an illustrative figure, Figure 22.

In this figure, we start at state s_0 and apply action **a**. Action **a** is not clear. It either takes us to state t_1 or to state r_1. We then apply more clear-cut actions **b**, which takes us to w_2 from t_1 and to u_2 from r_1 and then action **c** which takes us from w_2 to w_3 and from u_2 to u_3.

At time 1, we can either be at t_1 or at r_1. At time 3 we can either be at w_3 or at u_3.

Suppose at time 2 the nature of action **a** is clarified (publicly announced). It is clarified that t_1 is the correct successor. If we are at u_2 then we should be at w_2.

This is indicated by the double arrow (ρ relation $u_2 \twoheadrightarrow w_2$).

If we are at w_2, then we are at the correct place. This is indicated by the double arrows $w_2 \twoheadrightarrow t_1$. We now have several legitimate sequences of states:

1. $s_0 r_1 u_2 w_3 \ldots$

2. $s_0 t_1 w_2 w_3 \ldots$

We also have the tree relation sequences

3. $(s_0 r_1 u_2 u_3)$ (not a legitimate sequence)

4. $(s_0 t_1 w_2 w_3)$

We need connectives in the syntax that will give us complete control to describe the properties of the semantics and axomatise it.

The following is a possible choice.

1. Modality $\Box$ to follow the tree relation
   - $t \vDash \Box A$ iff for all s such that $tR_1 s$ we have $s \vDash A$
2. Yesterday operator for $\Box$
   - $t \vDash YA$ iff the R_1 predecessor of $t \vDash A$.
   - If $t = s_0$ then $YA = \bot$
3. Modality corresponding to ρ (i.e. the double arrow relation).
   - $t \vDash \boxminus A$ iff for all s such that $t\rho s$ we have $s \vDash A$.
4. Yesterday relation for $\boxminus$.
   - $t \vDash \mathbb{Y}A$ iff for the s such that $s\rho t$ we have $s \vDash A$ and if no such s exists then $\bot$.
5. Time constants D_n. $t \vDash D_n$ iff the distance of t from s_0 is n.

I think we have enough operators in the syntax to describe the semantics. We can write axioms and attempt to prove completeness.

It is not our purpose in this paper to put forward pure technical results, but we are giving enough details for the interested reader.

EXAMPLE 55. In Figure 21 we have $z < x < w$.

REMARK 56. To obtain a non-deterministic model we allow ρ in Definition 53 to be a general binary relation. To explain how this works consider Figure 23. The points y_1, y_2 and x are at distance m from t and they are all on different paths separating at z.

The points x and y_3 are also at distance m but they are on paths separating at t. So the path $s_0, \dots, z, \dots, u, \dots, t, s, \dots, x)$ made a choice of u over v_1 and v_2 and s over r.

If we let $x\rho y_3$ we are saying r is a correct non-deterministic choice and if we let $x\rho y_1 \wedge x\rho y_2$ we are saying that v_1 and v_2 are the right non-deterministic choice.

Note that by saying $x\rho y_3$ we are also implying that u is a correct non-deterministic choice, because only through u can we get to r.

However, we can also take the view that $x\rho y_3$ says that only as long as we can go through u, then r is the correct choice.

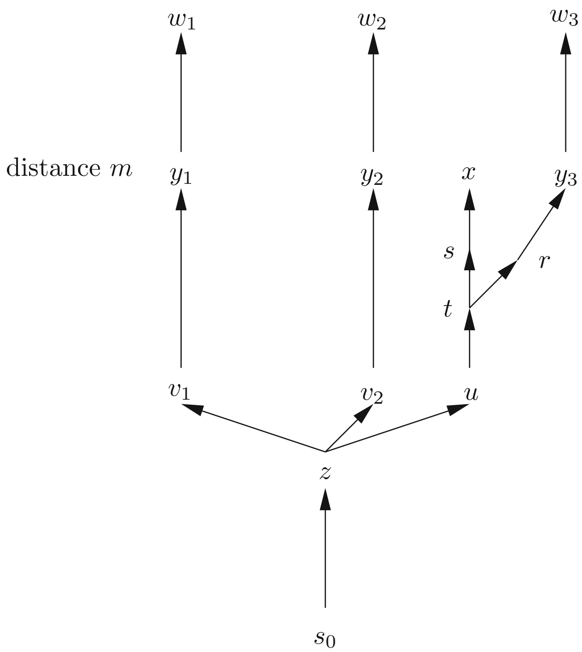


Figure 23. Non-deterministic model

4.8 Discussion and Comparison with traditional public announcement logic

REMARK 57. We begin by comparing Definitions 41 and 42 and Remark 43 with Definition 53.

Consider the following Figure 24. Ignore the double arrows, and consider only single arrows.

The arrows represent the relation R_1 for the set of nodes S.

Define the relation $\mathcal{R}$ (according to Definition 41) as follows:

$$\begin{aligned}
&t\ \mathcal{R}\ \{z_1, z_2\}\\
&z_1\ \mathcal{R}\ \{y, y_1\}\\
&z_1\ \mathcal{R}\ \{s_1, s_2, x\}\\
&x\ \mathcal{R}\ \{w\}\\
&y\ \mathcal{R}\ \{w'\}
\end{aligned}$$

We used the obvious recipe

$$(*) \qquad \alpha\ \mathcal{R}\ \{\beta | \alpha R_1 \beta\}$$

Consider the following public announcements in the sense of Definition 41:

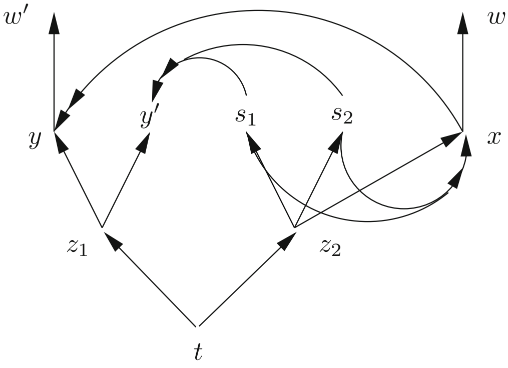


Figure 24.

(a_1) $(t, \{z_1, z_2\}, z_1)$

(a_2) $(z_1, \{y, y'\}, y)$

(a_3) $(z_2, \{s_1, s_2, x\}, x)$.

We now have a model frame in the sense of Definition 41. To turn this into a proper model we need to say what are the worlds and which worlds we assign to atoms. Let us do that. The worlds are the legitimate sequences, which are really the points in the graph, because the graph is a tree. This is in full agreement with Definition 42.

Now assume we live at the world s_1.

Our public announcements establish that the real sequence is (t, z_1, y, w'). So where does s_1 collapse to?

Obviously, if we have all the public announcements (a_1)–(a_3), then s_1 collapses to y. However, if we only have (a_1), then where does s_1 go to? (The other public announcements (a_2)–(a_3) may come later.) So our model lacks full information. This is what we have already remarked in Remark 43. Now consider the double arrows $s_1 \twoheadrightarrow y', s_2 \twoheadrightarrow y', x \twoheadrightarrow y$. These double arrows, when read according to Definition 53, all express the public announcement (a_1), but they also say more than that, they also say where s_1, s_2 and x are supposed to go to, in the event that public announcement (a_1) is put forward! There are provisions in Definition 53 (items 2.3 and 2.4) which ensure that this extra information does not contradict itself.

Furthermore, the double arrows also tell us at what time and place the public announcement is made. So the possible additional double arrow $s \twoheadrightarrow w'$ for example (which is not shown in Figure 24) not only tells us the public announcement (a_1) and where w is to go to (to w') but it also tells us that (a_1) was announced at time 4 (4 is the distance of w from the origin t.

REMARK 58. We now compare in more detail the Talmudic public announcement logic with the traditional public announcement logic as described in [40].

Assume we have modal operators $\Box_\varphi$ for φ a wff and $\Box_\alpha$ for α a non-deterministic public announcement statement in the sense of Definition 41. We have, in a model (S, R, t).

- $t \vDash \Box_\varphi A$ iff If $t \vDash \varphi$ (i.e. $t \in S_\varphi$) then $t \vDash A$ in the model $(S_\varphi, R \restriction S_\varphi, t)$ where $S_\varphi = \{x | x \vDash \varphi\}$.

  For comparison we have according to Definition 41

- $t \vDash \Box_\alpha A$ iff If t is a legitimate sequence in $(S, \mathcal{R}_\alpha)$ then $t \vDash A$ in $(S, \mathcal{R}_\alpha, t)$.

Both definitions move from a larger model (S, R) to a smaller model either made smaller by φ or by α.

So if we allow for any subset $T \subseteq S$ to be definable by a formula φ, the two definitions are the same.

In both cases the public announcement is metalevel and outside the model and we require that the point of evaluation t is not destroyed by the public announcement.

The refinement of Definition 53 is different

1. It is object level
2. It ties the announcment to a time and place.
3. It allows for the evaluation world to be destroyed by the announcement and displaces us and sends us to another world.
4. It has a backward effect in that it allows us to define $<$ as in Definition 53, item (3).

APPENDICES

APPENDIX A: A CONVENIENT VIEW OF MODAL LOGIC **K**

We view modal logic **K** as a time action logic, which is more convenient for our purpose.

4.9 Modal logic **K**

Modal logic **K** is formulated in the language of classical logic with the added unary operator $\Box$. The set of theorems is generated by the following axioms schemes:

1. All substitution instances of classical truth functional tautologies.
2. $\Box(A \to B) \to (\Box A \to \Box B)$
3. $\frac{\vdash A}{\vdash \Box A}$

The logic is complete for Kripke models of the form (T, R, Ω, h), where (T, R, Ω) is a tree with root Ω and h is an assignment giving for each atom q of the language a subset $h(q) \subseteq T$.

Note that actually modal **K** is complete for the class of finite trees.

Satisfaction is of the form $t \vDash A$, where $t \in T$ and A a wff and is defined recursively as follows:

- $t \vDash q$ iff $t \in h(q)$ for q atomic.
- $t \vDash A \wedge B$ iff $t \vDash A$ and $t \vDash B$
- $t \vDash \neg A$ iff $t \nvDash A$
- $t \vDash \Box A$ iff for all s such that s is an immediate successor of t in the tree we have that $s \vDash A$.
- A holds in the model (T, R, Ω, h) iff $\Omega \vDash A$.
- **K** is complete for this semantics in the sense that we have for any A:
  (*) A is a theorem of **K** iff A holds in every model.

4.10 Time action view of **K**

We now want to view this tree semantics for **K** as a time-action model. We first describe the tree (T, R, Ω) using successor functions. Let $\mathbb{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots\}$ be a set of unary successor functions capable of operating on Ω. We thus form all element sequences of the form

$$t = \Omega \mathbf{a}_1 \mathbf{a}_2 \mathbf{a}_3 \ldots \mathbf{a}_n$$

We define tRs to hold iff for some $\mathbf{a} \in \mathbb{R}$ we have $s = t\mathbf{a}$.

We call such models $\mathbf{m} = (\mathbb{R}, R, \Omega)$ *time action models*. We regard Ω as the initial state, the elements of $\mathbb{R}$ as actions $\mathbf{a} \in \mathbb{R}$ moving us from any state t to a new state $t\mathbf{a}$. Such a view is consistent with agent theory, if we regard as part of any state also the sequence of actions generating this state.

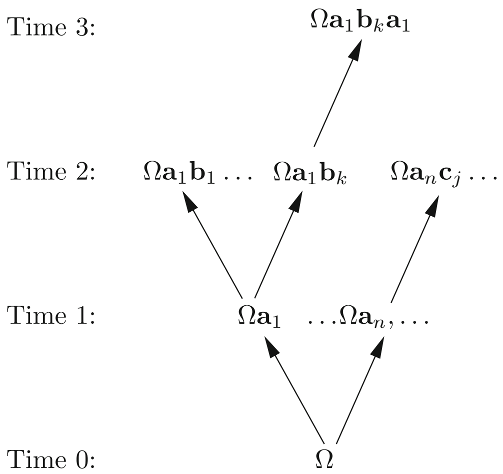
Figure 25 shows a time action model.

Time comes into the model if we take the view that time moves one unit when we perform any action.

So at time 0 we are at the initial state Ω.

If at time n we are at the state t and we apply action $\mathbf{a}$ then we move to time $(n + 1)$ and to state $t\mathbf{a}$.

Note the following


Time 3: $\Omega\mathbf{a}_1\mathbf{b}_k\mathbf{a}_1$
Time 2: $\Omega\mathbf{a}_1\mathbf{b}_1 \ldots$ $\Omega\mathbf{a}_1\mathbf{b}_k$ $\Omega\mathbf{a}_n\mathbf{c}_j \ldots$
Time 1: $\Omega\mathbf{a}_1$ $\ldots\Omega\mathbf{a}_n, \ldots$
Time 0: Ω

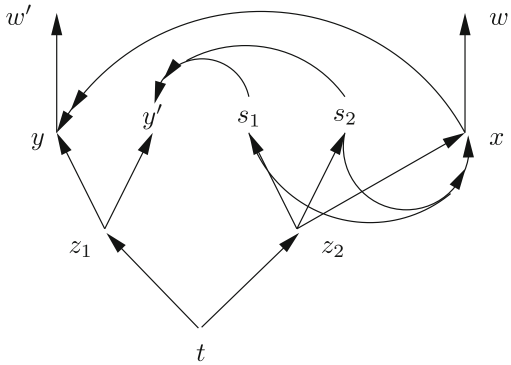

w'
w
y'
s_1
s_2
y
x
z_1
z_2
t

Figure 25.

(a) Actions $\mathbf{a}$ are deterministic and uniquely take a state t to the unique state $t\mathbf{a}$.

(b) The meaning of $t \vDash \Box A$ is that at state t any action $\mathbf{a}$ when applied will take us to a new state $t\mathbf{a}$ in which A holds.

(c) To do proper justice to this view, we need to specify the actions through their pre-conditions and post-conditions and present Ω as a compelte theory Δ_Ω defining the initial state. When we apply any action $\mathbf{a}$ to state Ω, we need a revision operator $*$ taking us from Δ_Ω to the theory of the new state $\Delta_{\Omega\mathbf{a}}$. Let $(\alpha_\mathbf{a}, \beta_\mathbf{a})$ be the precondition and post condition respectively of the action $\mathbf{a}$. Then we have

$$\begin{aligned}\Delta_{\Omega\mathbf{a}} &= \Delta_\Omega \text{ if } \Delta_\Omega \nvdash \alpha_\mathbf{a}\\ \Delta_{\Omega\mathbf{a}} &= \Delta_\Omega * \beta_\mathbf{a} \text{ if } \Delta_\Omega \vdash \alpha_\mathbf{a}\end{aligned}$$

The upshot of the above is that technically an assignment h into the model which is arbitrary in the traditional case of modal $\mathbf{K}$ corresponds to assigning $\Delta_\Omega, (\alpha_\mathbf{a}, \beta_\mathbf{a})$, and $*$ to the time action system.

Of course we can always take a tree model for traditional modal $\mathbf{K}$ and take the successor functions $\mathbf{r}_n$ and pretend they are actions and pretend that some $*$ exists such that for each t,

$$\Theta_{t\mathbf{r}_n} = \Theta_t * \beta_{\mathbf{r}_n}$$

where $\Theta_t = \{\varphi | \varphi \text{ does not contain } \Box \text{ and } t \vDash \varphi\}$.

If we do this we need the extra coherence condition that

$$\Theta_t = \Theta_s \Rightarrow \Theta_{t\mathbf{r}_n} = \Theta_{s\mathbf{r}_n}$$

which means in terms of the assignment h:

- for all atoms $q(t \vDash q \Leftrightarrow s \vDash q) \Rightarrow$ for all atoms $q(t\mathbf{r}_n \vDash q \Leftrightarrow s\mathbf{r}_n \vDash q)$.

The reader can read more about modal logic in [36] and generally about time action logic in [66]

APPENDIX B: PUBLIC ANNOUNCEMENT LOGIC

Modal jump operators

We begin with modal logic $\mathbf{K}$ with a jump operator. This is done as follows. We add to the language a unary operator JA. We add to the semantics of modal $\mathbf{K}$ a unary function $\mathbf{f}$ from $T \mapsto T$. So our models have the form $(T, R, \mathbf{f}, \Omega, h)$, where (T, R, Ω, h) is a tree model as before and $\mathbf{f}$ a unary function. We have

$$t \vDash JA \text{ iff } \mathbf{f}(t) \vDash A.$$

A well known jump operator in temporal logic is **Tomorrow**A.

Let $\mathbf{m}_i = (T_i, R_i, \Omega_i, h_i)$ be a family of models for $i \in I$. We can define a *fibring jump* operator through a function $\mathbf{g} : I \mapsto I$ as follows, $t \in T_i$

$$t \vDash_i JA \text{ iff } \Omega_{\mathbf{g}(i)} \vDash_{\mathbf{g}(i)} A.$$

In this statement we write $t \vDash_i A$ to mean satisfaction in modal $\mathbf{m}_i$.

The function $\mathbf{g}$ can be more general. We can assume T_i are all pairwise disjoint. So when we write $t \vDash A$, for $T \in \bigcup_i T_i$ then there is a unique i such that $t \in T_i$, and so we know to evaluate $t \vDash A$ as $t \vDash_i A$.

Now let $\mathbf{g} : \bigcup_i T_i \mapsto \bigcup_i T_i$. We can let

$$t \vDash JA \text{ iff } \mathbf{g}(t) \vDash A.$$

It is important to adopt the convenient point of view towards the operator J. We start with a model $\mathbf{m}$ and whenever we apply JA at a point t, we move to a new model $\mathbf{m}_{\mathbf{f}(t)}$. Thus the operator J moves us from one model to another.

4.11 Traditional public announcement logic

Such logics involve operators of the form J_φ, φ a wff, basically declaring publicly that φ is true. This is done in the context of interacting agents with knowledge operators and the public announcements are made by the agents to contribute to common knowledge.

Let $\mathbb{A}$ be a set of agents and for each $a \in \mathbb{A}$ let $\mathbb{K}_a$ be a knowledege operator.

This is a **K** modality satisfying the following additional axioms:

- $\mathbb{K}A \to A$, knowledge is true
- $\mathbb{K}A \to \mathbb{K}\mathbb{K}A$, positive introspection
- $\neg\mathbb{K}A \to \mathbb{K}\neg\mathbb{K}A$, negative introspection

For a detailed example involving the famous muddy children, see [79].

We start with an initial Kripke model for $\{\mathbb{K}_a | a \in \mathbb{A}\}$ of the form $\mathbf{m} = (S, R_a, h), a \in \mathbb{A}$ where S is the set of possible world and R_a are relations $R_a \subseteq S \times S$ satisfying the suitable conditions for $\mathbb{K}_a$ axioms. When we use the public announcement jump operator J_φ, φ a wff, we move to a model $\mathbf{m}_\varphi$ where

$$\mathbf{m}_\varphi = (S_\varphi, R_a^\varphi, h^\varphi)$$

where

$$S_\varphi = \{t \in S | t \vDash \varphi\}$$

$$R_a^\varphi = R_a \restriction S_\varphi$$

$$h^\varphi = h \restriction S_\varphi.$$

We have the condition

$$t \vDash_{\mathbf{m}} J_\varphi A \text{ iff } (t \vDash_{\mathbf{m}} \varphi \Rightarrow t \vDash_{\mathbf{m}_\varphi} A) \tag{$\sharp$}$$

This definition makes sense in the context of knowledge.

If φ is announced as true, we need consider only worlds in whcih φ holds. So we move from one model to another as more and more information is announced.

4.12 Public announcement logic for time-action modal K

How do we introduce the public announcement operators $J_\varphi A$ into our modal logic **K**, where we have the interpretation of time-action for the nodes $t = \Omega \mathbf{a}_1, \ldots, \mathbf{a}_n$?

We can start by adopting a technical approach, but this will not work. Let us see why. Start with a time action model

$$\mathbf{m} = (T, R, \Omega, h).$$

Let φ be any wff. Let $T_\varphi = \{t \in T | t \vDash \varphi\}$.

$$\mathbf{m}_\varphi = (T_\varphi, R \restriction T_\varphi, \Omega, h \restriction T_\varphi).$$

We can first try to define the semantical condition for the operator $J_\varphi A$, as follows. Assume $t = \Omega \mathbf{a}_1, \ldots, \mathbf{a}_n$. Define $t_0 = \Omega, t_{i+1} = t_i \mathbf{a}_{i+1}$ for $0 \leq i \leq n-1$. This makes $t = t_n$.

We let

$$t \vDash_{\mathbf{m}} J_\varphi A \text{ iff } \bigwedge_{i=0}^{n} t_i \vDash_{\mathbf{m}} \varphi \Rightarrow t \vDash \mathbf{m}_\varphi A. \tag{$\sharp$1}$$

The reasons for trying to use Condition ($\sharp$1) and not ($\sharp$) is as follows. We start with state Ω and apply actions $\mathbf{a}_1$ and $\mathbf{a}_2$ to get to $t = \Omega \mathbf{a}_1 \mathbf{a}_2$. If there is public announcement J_φ and $\Omega \mathbf{a}_1 \vDash \neg\varphi$, then even though we may have that $\Omega \mathbf{a}_1 \mathbf{a}_2 \vDash \varphi$, we could not have legitimately reached $t \mathbf{a}_1 \mathbf{a}_2$ in the first place. So our condition ($\sharp$1) is really conceptually the same as the traditional one ($\sharp$) given our time-action interpretation of modal **K**.

This attempt is not satisfactory. It is not fully compatible with the idea that a public announcement gives us more information about the model and becasue of that information we get a smaller model. Our time action model is generated from the initail state Ω (whose theory is Δ_Ω) via actions of the form $\mathbf{a} = (\alpha_{\mathbf{a}}, \beta_{\mathbf{a}})$ and a revision process $*$.

Thus

$$\begin{aligned}\Delta_{\Omega\mathbf{a}} &= \Delta_\Omega * \beta_\mathbf{a}\\ \Delta_{\Omega\mathbf{ab}} &= (\Delta_\Omega * \beta_\mathbf{a}) * \beta_\mathbf{b}\end{aligned}$$

The information that φ must be true must be in tune with the way the model is generated and therefore must affect the initial Δ_Ω, say we have

$$\Delta_\Omega^\varphi = \Delta_\Omega * \varphi$$

and must affect all subsequent action sequences, say recursively

$$\Delta_{t\mathbf{a}}^\varphi = \Delta_t^\varphi * (\beta_\mathbf{a} * \varphi).$$

This is one possible way of doing it. However, the Talmud is concerned with a different type of the need for public announcement. In the Talmud, we addres lack of clarity in the action $\mathbf{a}$ itself. Let $\mathbf{x}$ be an action variable where $\mathbf{x} \in E_\mathbf{x}$

$$E_\mathbf{x} = \{\mathbf{e}_1, \ldots, \mathbf{e}_k\}.$$

So we apply action $\mathbf{x}$ at t but we don't know whether we applied $\mathbf{e}_1$ or $\mathbf{e}_2, \ldots,$ or $\mathbf{e}_k$.

Thus after a sequence of further actions, say $\mathbf{b}_1, \ldots, \mathbf{b}_m$, we might be at any of the following k possible points, denoted by $\mathbf{s}(\mathbf{x}) = t\mathbf{x}\mathbf{b}_1, \ldots, \mathbf{b}_m$.

$$\left[\begin{array}{l} s_1 = t\mathbf{e}_1\mathbf{b}_1, \ldots, \mathbf{b}_m \\ s_k = t\mathbf{e}_k\mathbf{b}_1, \ldots, \mathbf{b}_m \end{array}\right.$$

A public announcement will have the form, say $J_{\mathbf{x}=\mathbf{e}_1}$ announcing at a later time that the action $\mathbf{x}$ taken in the past was $\mathbf{x} = \mathbf{e}_1$.

How such situations can arise and how to handle them is what discussed in this section.

5 THE HANDLING OF LOOPS IN TALMUDIC LOGIC, WITH APPLICATION TO ODD AND EVEN LOOPS IN ARGUMENTATION

Background

The Talmud is a body of arguments and discussions about all aspects of the human agent's social, legal and religious life. It was completed over 1500 years ago and its argumentation and debates contain many logical principles and examples very much relevant to today's research in logic, artificial intelligence, law and argumentation.

In a series of books on Talmudic Logic, the authors have studied the logical principles involved in the Talmud, one by one, devoting a volume to each major principle

We have just finished writing Volume 5, entitled *Resolution of Conflicts and Normative Loops in the Talmud*, and the present section describes how the Talmud deals with even and odd loops and compares the results with open issues in argumentation.

For other English papers corresponding to previous sections, see [3; 10; 5; 14; 5; 7] and of course earlier sections of this chapter which make use of them.

We start by looking at two typical loops, as in Figures 26 and 27.

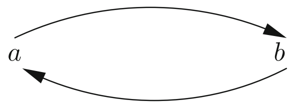


Figure 26.

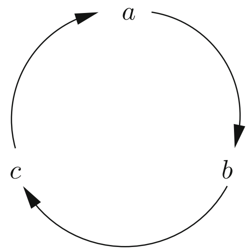


Figure 27.

We need to give some definitions.

An abstract network has the form (S, R), where S is a set of abstract nodes (arguments) and $R \subseteq S^2$ is the attack relation. Traditional research looks at extensions, these are subsets of S satisfying certain conditions (formulated in terms of R). Given (S, R) there may be several possible extensions of several types. In our case, for example, Figure 26 has three complete extensions $\{a\}, \{b\}$ and $\varnothing$, and Figure 27 has only one extension $\varnothing$.

Current research in argumentation, which relates to such loops and which connects with Talmudic logic, has two aspects:

1. Giving new definitions of extensions which can apply to abstract argumentation networks containing loops and allow us to get some new extensions other than "all undecided".

2. Adding extra information to the argumentation network which helps resolve the loops or help choose an extension.

The extra information one can add to the nodes of the network can be valuations or preferences among nodes. Mathematically one can look at valuations only, as preferences can be derived from them.

When we add valuations, we add a function $V : S \mapsto U$ where U is a value domain, giving some value to each $x \in S$.

V can be used in two extreme ways:

(a) Use V in the definition of extensions, by modifying the network or by disregarding and removing attacks, etc.

(b) Calculate the extensions without using V (i.e. ignoring V) and then using V to choose one's favourite extension or modify existing extensions and create new modified extensions.

(c) There is a third way, highly recommended by some members of the community, which is to use V in combination with the internal structure of the argument. (Note that V is not defined on arguments here but on components of arguments).

(a) is supported by Leila Amgoud and Trevor Bench-Capon.

(b) is supported by the 1500 years old Talmudic logic and recently by a 2010 paper by Toshiko Wakaki, [108].

(c) is supported by Henry Prakken in a 2010 paper [95].

The (b) and (c) approaches maintain consistency while (a) is problematic. See a critique by Martin Caminada [34]. We are grateful to Martin Caminada for providing us with the above information, as well as sending us his critique of approach (a).

Our plan for this section is very simple. In Section 2 we present the notion of Shkop extension to an abstract network (S, R) and compare it with Baroni's and his colleagues [20; 21] CF2 extensions.[79]

In Subsection 3 we discuss some counter examples by Martin Caminada. In Subsecection 4 we conclude this section. In a follow-up paper, yet to be written, we give examples of how the Talmud offers valuations to resolve loops of odd and even types and how the Talmud chooses extensions.

5.1 Shkop extensions

We begin with a motivating Talmudic example, the dates are all in the same year, say 2010.

EXAMPLE 59 (The divorce). Jane is married to John. She develops some feelings for Frank and wants a divorce from John. Frank is a rich man and promises to compensate John generously if he cooperates. We now have the following temporal sequence:

[79] Rabbi Shimon Shkop, 1860–1930. A Talmudic scholar analysing many logical principles in the Talmud.

Jan 01: John gives divorce papers to Jane. The divorce is conditional on Jane marrying Frank by the 31st of March. Such conditional divorces are allowed in the Talmud. If Jane marries Frank before 31st March then all is well. If Jane does not marry Frank by 31st March then the divorce papers, the beginning, from January 01 are nullified and the divorce is not valid from Jan 01. This is Talmudic legal backwards causality.

Feb 01: Jane takes her divorce papers and marries Terry. This marriage is valid because Jane's divorce papers are valid. Jane can still potentially fulfil the condition mentioned in the divorce papers; she can still divorce Tery and marry Frank.

31st March: Jane, without getting a divorce from Terry, goes and marries Frank.

There is no doubt that Jane is a naughty girl! Frank is a bit paranoid, asking John to give Jane a conditional divorce.

Now we seem to have landed in a logical loop.

Let us build up an argumentation network based on this story.

The base logic is classical temporal logic. The base theory in the logic is the following:

1. If x is married to someone then x cannot marry someone else.
2. If x is married to y at time t then x continues to be married to y until there is a divorce or death.
3. (a) A divorce can be given at time t, conditional on an action taken at time $s > t$.
   (b) If the action is not taken at time s then there is backward causality and the divorce is not valid from time t.
   (c) If the action is taken at time s then the divorce is valid at time t.
   (d) At any time $t', t \leq t' < s$, the divorce given at time t on a condition to be fulfilled at time s, is considered valid at time t' as long as there is the reasonable possibility, as seen from time t', that the condition will be fulfilled at time s.
4. **Fact**: John gave a divorce to Jane on January 01, conditional on Jane marrying Frank by March 31.
5. **Fact**: Jane married Terry on Feb 01.
6. **Fact**: Jane married Frank on March 31, without ever getting a divorce from Terry.

7. Note: It is possible for x to give a divorce to y at time t on the condition that y marries z ($z \neq x$) at time $s > t$.

   One might argue that at time s, we have a problem:

   y is still married to x therefore y cannot marry z. It is only when y marries z at time s that y is divorced from x at time t and is therefore able to marry z at time s.

   Since we allow for such conditions, we regard marrying z at s and enabling the divorce at t as simultaneous.

   The answer is that the condition of marrying z is not an enabling condition for the divorce papers but a nullifying condition. If it is not fulfilled the divorce papers are nullified.

We now consider the following arguments seen from the temporal point of view of March 31.

DJJ- John's divorce from Jane on January 01 is not valid.
The reasoning in this argument from base data goes as follows:

On February 01, the divorce was valid because there was the possibility of fulfilling the condition of the divorce, from Rule (3d). Therefore the marriage to Terry (Fact (5)) is valid and does not nullify the divorce, since Jane can still divorce Terry and marry Frank (Rule (3d)).

Therefore at the time March 31, when Jane married Frank without divorcing Terry, her marriage to Frank was not valid (Rules (1) and (2)). Hence, since the condition of the divorce was not fulfilled, the divorce is not valid.

MJT+ Jane's marriage to Terry on Feb 01 is valid.
The argument goes as follows:

Since Jane got a conditional divorce from John and the condition can still be fulfilled her divorce stands and she can marry Terry.

MJF+ Jane's marriage to Frank is valid.
the argument for that is as follows:

Assume the marriage to Frank is not valid. Then Jane's divorce from John is not valid. Hence her marriage to Terry is not valid. But then Jane has a conditional divorce from John and she is not married to Terry, therefore she is free to marry Frank and the marriage is valid. Therefore since $\neg$**MJF+** $\rightarrow$ **MJF+**, we conclude **MJF+**.

We now get the argumentation loop presented in Figure 28.

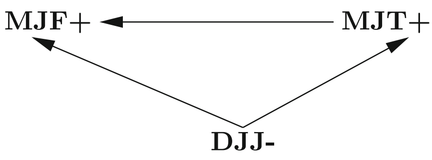


Figure 28.

It is clear that we have an odd loop here and the only Dung extension is $\varnothing$, being all undecided. However, life must go on and we need a resolution as to whom Jane is married to! Is she married to John, to Terry or to Frank?!

Here we introduce the intuitive Rabbi Shkop principle:

Shkop principle

If by assuming $x =$ in, we deduce that $x =$ out, then surely x must be out.

Let us apply this to our example. We have three possibilities for the choice of x, see Figure 28.

1. $x =$ **DJJ**$-$
2. $x =$ **MJT**$+$
3. $x =$ **MJF**$+$

We reason against the direction of the attack arrows. This reasoning is done later on, see Example 64 below for the calculation.

We get three extensions for each one of the choices of x:

1. Marriage to Frank is valid.
2. Divorce not valid — Jane is married to John.
3. Marriage to Terry is valid.

Commonsense dictates that we should not test the validity of the divorce because at the time (and here we make use of the temporal sequence) we did not know what Jane was going to do. Similarly we should not test the validity of the marriage to Terry because Jane could still have divorced him. So the only test is that of the validity of marriage to Frank. This test gives by the Shkop principle that **MJF**$+$ $=$ out and therefore the network looks like Figure 29, (see also Example 64 below for a detailed analysis).

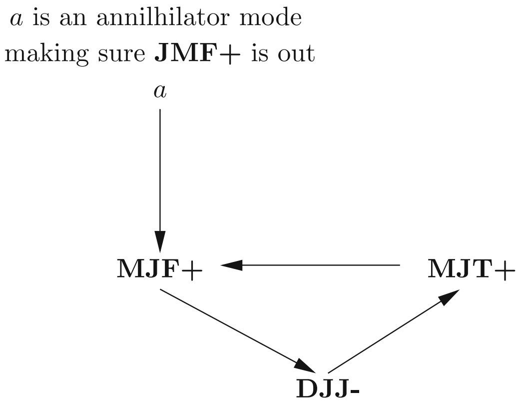


Figure 29.

Figure 29 has the extension:

a	=	in
DJJ−	=	in
MJT+	=	out
MJF+	=	out

In the above considerations we kept the temporal aspects in the metalevel. We can include these aspects in the object level. We time stamp each argument and each attack arrow, according to the way the story unfolds. If we do this we get Figure 30.

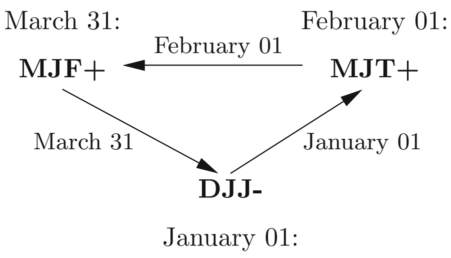


Figure 30.

Obviously the loop occurs on March 31. So we have to do the Shkop test on the March 31 argument, which is **MJF+**.

In general we can talk about Shkop temporal argumentation frames of the form $\mathbf{N} = (S, R, \mathbf{T})$, where (S, R) is an ordinary network and $\mathbf{T}$ is a time stamping function:

$$\mathbf{T} : S \cup R \mapsto \text{ Time axis.}$$

For any choice of time t we look at the network

$$\mathbf{N}_t = (S_t, R_t),$$

where

$$\begin{array}{l} S_t = \{a \in S | \mathbf{T}(a) \leq t\} \\ R_t = \{(x, y) \in R | \mathbf{T}(x, y) \leq t\} \end{array}$$

Given $a \in S$ with $\mathbf{T}(a) = s$, we check according to Shkop the test $a = 1$? in the network $\mathbf{N}_s = (S_s, R_s)$.

Let us now be a bit more formal about Shkop extensions. Our aim is to offer the argumentation community the notion of Shkop semantics, and compare it with CF2 or Stage semantics. To do that, we need to generalise the intuitive Shkop principle in a sensible way.

For reasons of clear exposition, we find it advantageous to actually start from a recent paper of Martin Caminada, entitled Preferred semantics as Socratic discussion [32].

Caminada sets himself to give a game theoretic answer to the question:

Q: Given (S, R) and $a \in S$, can a be an element of some admissible extension?

His method is to assume that $a =$ in and see by Socratic discussion whether such a position can be maintained. The method is best explained by two examples.[80]

EXAMPLE 60. Consider Figure 31

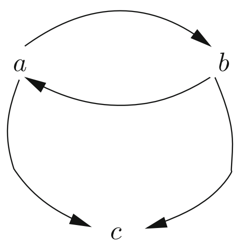


Figure 31.

We ask can we have $c =$ in in some extension? We proceed as follows:

1. $c =$ in, assumption

[80]Appendix A offers a Tableaux algorithm for this test.

2. $b =$ out, from (1)

3. $a =$ in, from (2)

4. $a =$ out, from (1)

We get a contradiction. The assumption $c =$ in, lead us, using the attack rules and the geometry of the figure, that both $a =$ out, and $a =$ in.

Thus the answer the question about c is that it cannot be in, it must be out.

EXAMPLE 61. Consider Figure 27. Ask the question can $c =$ in? Let us check:

1. $c =$ in, assumption

2. $b =$ out, from (1)

3. $a =$ in, from (2)

4. $c =$ out, from (3)

So again the answer is negative, there is no extension in which $c =$ in.

REMARK 62. Note that the proofs in the Caminada Socratic discussion obtain a contradiction by using the direction in the graph against the arrow. Thus if we have

$$x \to y \to z$$

and we assume $y =$ in, Caminada is allowed to deduce $x =$ out, going against the arrow, but is not allowed to deduce $z =$ out going with the arrow.

It seems that even with this restriction, the Socratic discussion is strong enough to identify all nodes a in the network for which $a =$ in is impossible.

Caminada's paper stops when we get our answers to the question of whether $a =$ in is possible or not.

Now let us use these two examples to explain what Shkop does. Shkop introduced a principle for resolving loops:

Shkop's original principle

If the test assumption $a =$ in leads to the conclusion that $a =$ out, then a must be annihilated and be out.

To implement such a principle we need some notation. Let (S, R) be a network and let the elements of S be denoted by lower case letters. Let us add for any $a \in S$ a new annihilator letter, capital A.

With the above notation, let us redo Examples 60 and 61 according to Shkop.

EXAMPLE 63 (Doing Example 60 according to Shkop). We start by testing $c =$ in in Figure 31.

1. $c =$ in, test assumption
2. $b =$ out, from (1)
3. $a =$ in, from (2)
   The Caminada Socratic discussion goes against the arrow and would continue

4*. $a =$ out, from (1), a contradiction, because we get both $a =$ in and $a =$ out.
   The Shkop original principle requires us to get $c =$ out for a contradiction, because our original test was for $c =$ in?. Therefore we need to go forward with the arrow using (3), as this is the only way to get back to c, and get (4) below. Going forward:

4. $c =$ out, from (3)
5. From (1)–(4) we get that c must be annihilated by the Shkop principle.

This means that we replace Figure 31 by Figure 32.

We may now feel comfortable, allowing ourselves to go both backwards and forwards with the arrow, and thus maintaining the intuitive spirit of the Shkop principle. This, however, is problematic. Caminada has shown a counter example which is problematic. We discuss this later in Section 3. So we cannot allow ourselves to prove forward with the arrow. So we need to modify the Shkop principle.

Our choice of modifying the Shkop principle is to state:

- If $a =$ in leads to a contradiction then a must be out. In deriving the contradiction, we use reasoning going backwards with the arrow only, see Appendix A. Once the contradiction is derived we introduce an annihilator for a.

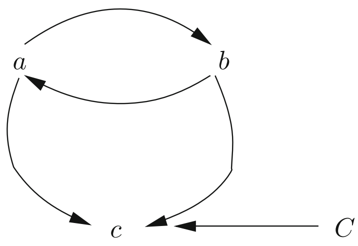


Figure 32.

Coming back to the argumentation network of Figure 31, having tested $c =$ in?, we can continue to test $a =$ in and test $b =$ in but this will not require any more annihilators.

The Shkop extensions for Figure 31 are obtained by taking ordinary extensions for Figure 32 and ignoring the annihilators. In the case of Figure 31 the Shkop procedure made no difference but for Figure 27 it does as it resolves loops.

EXAMPLE 64 (Doing Example 61 according to Shkop). We have three tests to conduct:

Test 1: $c = \text{in}$

Test 2: $b = \text{in}$

Test 3: $a = \text{in}$

Test 1

1. $c = \text{in}$, test assumption
2. $b = \text{out}$, from (1)
3. $a = \text{in}$, from (2)
4. $c = \text{out}$, from (3)
5. Using the Shkop principle c must be annihilated and Figure 27 replaced by Figure 33.

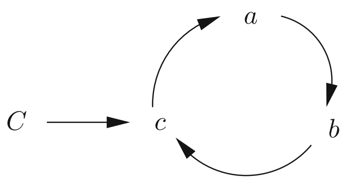


Figure 33.

Figure 33 is a new network and we can apply the Shkop test to it. We will get no more contradictions. The Dung extension for it is $\{C, a\}$.

The other tests will give us Figures 34 and 35.

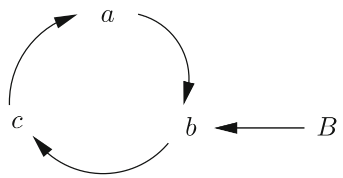


Figure 34.

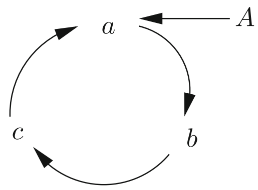


Figure 35.

The normal Dung extensions for Figures 34 and 35 are respectively $\{B, c\}$ and $\{A, b\}$.

According to Shkop, the Shkop extensions for Figure 27 are obtained from the normal extensions of Figures 33, 34 and 35 by ignoring the annihilators letters.

Thus we get the extensions $\{a\}, \{b\}, \{c\}$. Notice that these are the conflict free sets of Figure 27.

We ask the reader to remember this because we shall compare the Shkop extensions with Baroni's CF2 extensions.

REMARK 65. The reader should note that the Shkop procedure was originally intended for elements x of a network which are part of an odd loop, see Example 59. Once the element is found to be out by the Shkop principle, we move to a new network containing the annihilator X of x and we deal with the new network only. Shkop would never test $c = \text{in}$? immediately (at that moment, if we take into account the temporal aspect, see Section 4) in Figure 26 because c is not part of a loop. He would test $a = \text{in}$? and $b = \text{in}$? and find no contradiction. For the sake of mathematical completeness and generalising Shkop, we can allow the use of the Shkop principle to any x in the network. The test is similar to the Caminada Socratic discussion (see Appendix A for a Tableaux algorithm doing the same as Caminada's Socratic discussion), and if $x = \text{in}$ is found contradictory, this means that x must be out. Thus adding the annihilator X with $X \to x$ to the network will make no difference and we get an equivalent network.

We therefore put forward the Generalised Shkop Principle:

Generalised Shkop Principle

Let (S, R) be a network and let $a \in S$. If the assumption $a = \text{in}$ leads to a contradiction (i.e. for some $x \in S$, we get both $x = \text{in}$, and $x = \text{out}$) by reasoning only backward against the direction of the arrow (as Caminada does in his Socratic discussion, or as we do in Appendix A using Tableaux) then a must be out. To ensure that a is out, we move to a new network $(S \cup \{A\}, R \cup \{A, a)\})$, where A is a new letter, being the annihilator of a.

Note that Caminada proved in his Socratic paper that for a network (S, R) and $a \in S$ the condition:

- The assumption $a =$ in leads to a contradiction by correctly reasoning backwards against the direction of the arrow.

is equivalent to the declarative condition

- a is not a member of any admissible set.

We can therefore formulate the Generalised Shkop principle in an equivalent declarative way as follows:

Generalised Shkop principle (declarative)

Let (S, R) be a network and let $a \in S$. If a is not a member of any admissible set, then a must be out. To ensure that a is visibly out we move to a new network $(S \cup \{A\}, R \cup \{(A, a)\})$, where A is a new letter, being the annihilator of a.[81]

We are now ready to define the notion of Shkop extensions.

DEFINITION 66.

1. Let $\mathbf{N} = (S, R)$ be a finite argumentation network. Assume elements $y \in S$ are denoted by lower case letters. For each such y let Y be the annihilator of y.

   We define by induction the notions of

   (a) $(y_1, \ldots, y_k), y_i \in S$ is a legitimate Shkop sequence.

   (b) $\mathbf{N}_{(y_1,\ldots,y_k)}$ is a Shkop model dependent on $(y_1, \ldots, y_k)$.

   Case $k = 1$
   y_1 is a legitimate Shkop sequence if y is not a member of any admissible set of (S, R) (or equivalently by Caminada [32], if the assumption $y =$ in, in (S, R) leads to a contradiction using Caminada Socratic discussion). In this case let

$$\begin{aligned} \mathbf{N}_{y_1} &= (S \cup \{Y_1\}, R \cup \{(Y_1, y_1)\}) \\ &= (S_{y_1}, R_{y_1}). \end{aligned}$$

   Case $k + 1$
   Assume $(y_1, \ldots, y_k)$ is a legitimate sequence and assume $\mathbf{N}_{(y_1,\ldots,y_k)}$ is well defined. Let $y_{k+1} \in S$ be a point such that y_{k+1} is different from all $y_1, \ldots, y_k$. Assume that y_{k+1} is not a member of any admissible set in $\mathbf{N}_{(y_1,\ldots,y_k)}$, (or equivalently the assumption $y_{k+1} =$ in, in

[81] So for example the Liar paradox network $(\{a\}, \{a \to a\})$ becomes the network $(\{A, a\}, \{a \to a, A \to a\})$.

the network $\mathbf{N}_{(y_1,\ldots,y_k)}$ leads to a contradiction using Caminada Socratic discussion). Then $(y_1,\ldots,y_{k+1})$ is a legitimate sequence and let $\mathbf{N}_{(y_1,\ldots,y_{k+1})}$ be $(S_{(y_1,\ldots,y_{k+1})}, R_{(y_1,\ldots,y_{k+1})}$, where

$$\begin{array}{rcl} S_{(y_1,\ldots,y_{k+1})} & = & S_{(y_1,\ldots,y_k)} \cup \{Y_{k+1}\} \\ R_{(y_1,\ldots,y_{k+1})} & = & R_{(y_1,\ldots,y_k)} \cup \{(Y_{k+1}, y_{k+1})\}. \end{array}$$

2. Let $(y_1,\ldots,y_k)$ be a legitimate sequence. Let n be the number of elements of S. Then we say the rank of $\mathbf{N}_{(y_1,\ldots,y_k)}$ is $n-k$.

3. Let $(y_1,\ldots,y_k)$ be a legitimate sequence. Let $\mathbf{N}_{(y_1,\ldots,y_k)}$ e its associated Shkop network. We say $\mathbf{N}_{(y_1,\ldots,y_k)}$ or equally $(y_1,\ldots,y_k)$ is *clean* iff there are no legitimate sequences extending $(y_1,\ldots,y_k)$. Alternatively, iff for any $y \in S, y \neq y_i, i = 1,\ldots,k$, we have that the test $y =$ in does *not* lead to a contradiction.

4. Let $\mathbf{N}_{(y_1,\ldots,y_k)}$ be clean. Then we define the set of Shkop extensions of $\mathbf{N} = (S, R)$ as derived from $(y_1,\ldots,y_k)$.
   Notation
   $$\mathbb{E}^{\text{Shkop}}_{(y_1,\ldots,y_n)}$$
   to be defined as follows.

   Let E be any ordinary Dung extension of $\mathbf{N}_{(y_1,\ldots,y_k)}$ or equivalently let λ be any Caminada labelling for $\mathbf{N}_{(y_1,\ldots,y_k)}$, then $E \cap S$ (or equivalently) $\lambda \restriction S$ be an element of $\mathbb{E}^{\text{Shkop}}_{(y_1,\ldots,y_k)}$.

5. We now define the notion of all Shkop extensions of a finite network $\mathbf{N} = (S, R)$. We define the set of all Shkop extensions of $\mathbf{N}$ to be

$$\mathbb{E}^{\text{Shkop}}_{\mathbf{N}} = \bigcup_{\substack{(y_1,\ldots,y_k) \\ \text{clean}}} \mathbb{E}^{\text{Shkop}}_{y_1,\ldots,y_k)}$$

REMARK 67. Note that this is our definition based on the generalised Shkop principle. We can give restricted variations of it. For example, following Baroni *et al.* in their paper [21] of SCC recursiveness, we can first rewrite (S, R) as an acyclic ordering of maximal loops and then apply the Shkop procedure to loop elements starting from the top loops. This is like the way the CF2 extensions are calculated. We shall give a substantial example below to show you what happens.

It is now time to give some more Shkop examples.

EXAMPLE 68. Consider Figure 36

Testing b and then testing a or testing a and then testing b will lead to the same Figure 37.

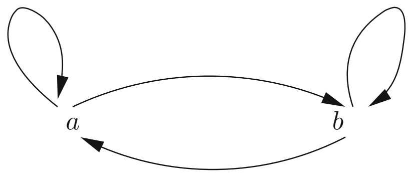


Figure 36.

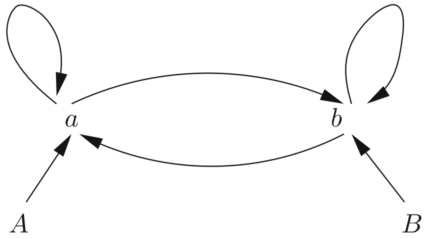


Figure 37.

Therefore the Shkop extension of Figure 36 is $\{a = \text{out}, b = \text{out}\}$.

This does not contradict the usual Dung extension of all undecided!

EXAMPLE 69 (Shkop compared with CF2). Consider the network in Figure 38. This figure appears in [69] as an example of how Baroni's CF2 semantics works. Gaggl and Woltran have a program which can compute the CF2 extensions.

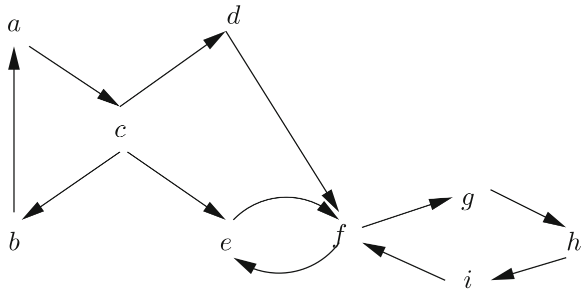


Figure 38.

The CF2 extensions for Figure 38 are the following:

$$\begin{aligned} E_1 &= \{c, f, h\} \\ E_2 &= \{c, g, i\} \\ E_3 &= \{b, d, e, g, i\} \\ E_4 &= \{a, d, e, g, i\}. \end{aligned}$$

The CF2 semantics would start with the top cycle $\{a, b, c\}$. They would take maximal conflict free subsets which are in this case $\{a\}, \{b\}, \{c\}$ and then arbitrarily decide on the three assignment:

1. $c =$ in, $b =$ out, $a =$ out
2. $b =$ in, $c =$ out, $a =$ out
3. $a =$ in, $c =$ out, $a =$ out

Having now given values to a, b and c, one can propagate the values to the rest of the network and get extensions.

For example:

If $c =$ in, then $d = e =$ out.

Therefore $f =$ in and hence $g =$ out, $h =$ in and $i =$ out.

We got ourselves an extension by breaking the loop $\{a, b, c\}$. The alternative, if we follow traditional Dung style approach is to have one extension only = all undecided.

The method makes sense, it is not arbitrary, it is not just a technical device to generate extensions.

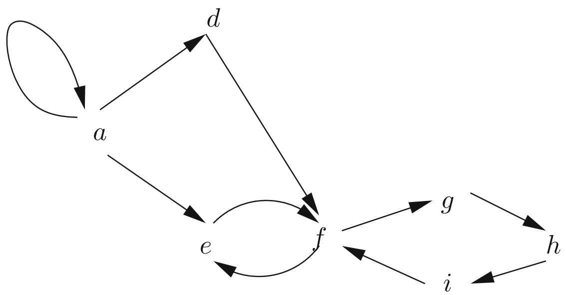


Figure 39.

CF2 would take maximal conflict free subsets of the loop $\{a\}$, which is the empty set, therefore d and e are in and so f is out, g is in, h is hout and i is in.

Now let us look at Shkop extensions of Figure 38.

Option 1

Accept the procedure where we start from the top loops. Call this top-down Shkop procedure. In this case we start from $\{a, b, c\}$ and ask, as in Example 64,

Test 1: $a =$ in
Test 2: $b =$ in
Test 3: $c =$ in

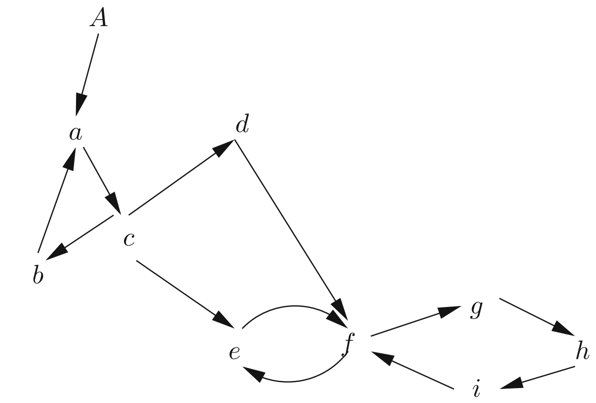


Figure 40.

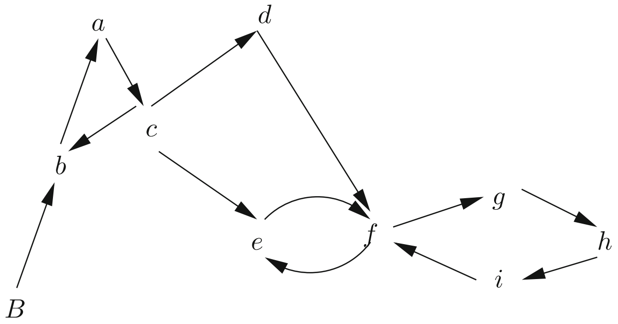


Figure 41.

This will yield Shkop figures 40, 41 and 42.

From Figure 40 we get the extensions E_1 and E_2. From Figure 41 we get Extension E_4 and from Figure 42 we get extension E_3.

In the case of Figure 39, using the Shkop procedure on $a = \text{in}$ will give Figure 43.

and we get the extension $\{d, e, g, i\}$.

Let us now check what happens if we allow the Shkop process to start from any point. Let us start with $d = \text{in}?$ and then $e = \text{in}?$. We will get that both need to be annihalated. If we carry on asking $a = \text{in}?$ or $b = \text{in}?$

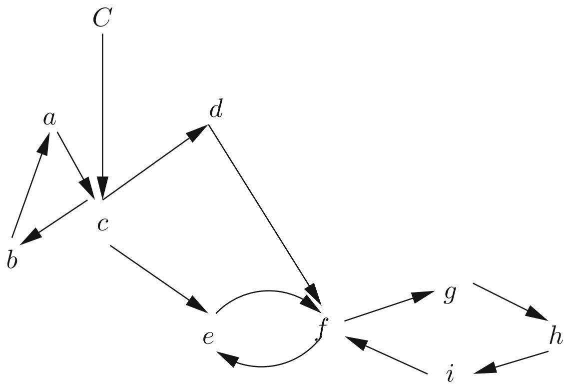


Figure 42.

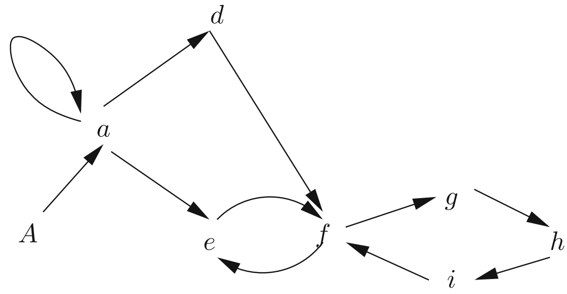


Figure 43.

or $c =$ in? we get the extensions

$$\begin{array}{c}\{c, f, h\}\\ \{c, g, i\}\\ \{a, f, h\}\\ \{a, g, i\}\\ \{b, f, h\}\\ \{b, g, i\}\end{array}$$

This section is mainly qualitative. A more mathematical exposition will need to address some open problems.

Problem 1
Under what circumstances is the top down Shkop process the same as CF2?

Problem 2
Is there a set of equations in the equational approach [67] characterising say the top down Shkop extensions?

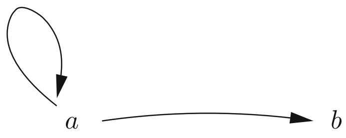


Figure 44.

Note that all extensions obtained by the Shkop procedure are stable. There is no undecided. Shkop kills all undecided!

REMARK 70. The reader may seek some meaning to the Shkop algorithm. For this, see the conclusion Section 4. The reader must remember that Talmudic logical argumentation and debate was conducted from the first to the end of the fifth centuries and was used in Jewish communities in the world during the following 1500 years.

Rabbi Shkop just explained the principles involved and we in this section are formally modelling them in terms of known abstract argumentation methods.

The principle works!

REMARK 71 (Comparison with stage semantics). Stage semantics is discussed in detail in [31]. It has similarities with the Shkop extensions but it is not the same. Both ignore self loops but stage semantics may ignore arguments which are not attacked by any other argument. Shkop extensions never do that.

We take our examples from [31]. Consider Figure 44

Stage semantics will ignore the self looping a and will have the extension $\{b\}$. The same is the case with the Shkop semantics. They both agree on $b =$ in. Stage will say $a =$ undecided while Shkop will say $a =$ out.

As a second example from [31], take Figure 45.[82] Shkop will not agree here with the traditional extension. According to Shkop we have

$$a = \text{in},\ b = \text{out},\ c = \text{out}.$$

The traditional extension will have

$$a = \text{in},\ b = \text{out},\ c = \text{undecided}.$$

Stage semantics allows for two extensions: the first one is the same as the traditional one

$$a = \text{in},\ b = \text{out},\ c = \text{undecided}.$$

The second one is

[82] In fact Pietro Baroni and Massimilano Giacomin invented this figure in order to show that CF2 semantics has some advantages above stage semantics.

$a =$ undecided, $b =$ in, $c =$ out.

Note that in the second stage a is not in, even though it has no attackers. This is rather strange. Caminada has proved, however, that every argumentation network has at least one stage extension which contains its ground extension. So it can be well behaved. Compare the stage semantics result for the network of Figure 45 with Example 72 and the considerations leading to Figure 46. We get the stage semantics if we go forward. Is this a coincidence? We think it is.

5.2 Caminada counter examples: A discussion

Martin Caminada read an earlier version of Section 2 and gave us penetrating comments and devastating counter examples. The aim of this Section is to put forward an alternative formulation of the Shkop principle which maintains the spirit of Shkop while avoiding the counter examples of Caminada.

We need to summarise the intellectual chain of reasoning events.

(1) The original Shkop principle, as formulated by Shkop, says as follows:

(*1) Let $\mathbf{N} = (S, R)$ be a network. Let $x \in S$. Assume (test) $x =$ in. If one can prove that this entails $x =$ out, then surely x must be out.

Our modelling of this principle was to move to the network $\mathbf{N}_x$, as defined in Definition 66.

Shkop does not specify what it means "to be able to prove that $x =$ in entails $x =$ out". We adopted the Caminada Socratic method to give meaning to this notion.

(2) Here we had a problem. Caminada's method uses reasoning against the direction of the arrow. So if we have, for example

$$y \to x \to z$$

and we test the assumption $x = 1$, then Caminada allows us to deduce $y = 0$, but we are not allowed to deduce $z = 0$.

The difficulty with this is that Shkop formulated his principle by saying "$x =$ in can prove $x =$ out".

It is the same x.

The "same x" restriction is OK for cases of pure loops of the form

$$x \to a_1 \to a_2 \to \ldots \to a_k \to x$$

We can prove $x =$ in implies $x =$ out by going backwards, but for cases like Figure 31 (the test case assuming $c =$ in) we cannot get $c =$ out by going against the arrow only, as discussed in Example 63.

Our original modification of Shkop principle was to allow forward reasoning with the arrow. However, Caminada landed a devastating counter example on this attempt (see Example 72 below).

We therefore reformulated the generalised Shkop principle in a safe way, as follows.

(*2) Let $\mathbf{N} = (S, R)$ be a network. Let $x \in S$. Assume (test) $x =$ in. If one can prove a contradiction from this assumption, say that for some $y \in S$, both $y =$ in and $y =$ out are derivable, then surely x must be out, and move to $\mathbf{N}_x$

The above is equivalent to the following (in view of Caminada's Socratic paper).

(*3) Let $\mathbf{N} = (S, R)$ and let $x \in S$. If x is not part of any extension (equivalently if there is no Caminada labelling λ with $\lambda(x) =$ in), then surely x must be out and we move to $\mathbf{N}_x$

So the Shkop extensions and Shkop semantics are obtained by systematically annihilating all points which cannot be part of an extension, as defined in Definition 66. This is a Draconian instrument. Note that it needs to be done in sequence, one node at a time.

EXAMPLE 72 (Caminada's counter example). Consider Figure 45

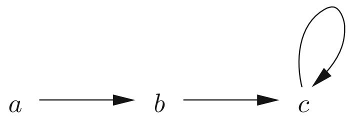


Figure 45.

Let us test $a = 1$? allowing reasoning both with and against the arrow. We reproduce Caminada's reasoning

1. $a =$ in, assumption
2. $b =$ out, from (1)
3. We now do case analysis for c.

   Case 3a $c =$ in
   In this case we continue

(4a) $c =$ out, since c attacks itself.

(5a) $b =$ in, from (4a)

(6a) $a =$ out, from (5a)

Case 3b $c =$ out

(4b) $b =$ in, since $c =$ out

(5b) $a =$ out, from (4b)

4. Since in both cases we get $a =$ out, then by the Shkop principle surely $a =$ out and we move to Figure 46.

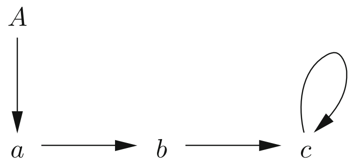


Figure 46.

Clearly this is not acceptable.

Later on we shall modify the forward proof procedures by means of Labelled Deductive Systems and hopefully avoid the Caminada counter example.

We shall now show the idea behind this modification. Let us do the proof again, using our idea:

1. $a =$ in, assumption

2. $b =$ out, from (1)

3. we now have a node c which is not attacked by any node which is in, and instead of doing a case analysis, let us ask, by way of a subcomputation, can $c =$ in?

Subcomputation

- Given assumptions: $a =$ in, $b =$ out

- we test: $c =$ in.

(3.1) $c =$ in, assumption

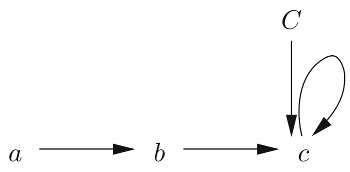


Figure 47.

(3.2) $c =$ out, from (3.1)

Therefore using the Shklop principle c surely must be out, and we move to Figure 47.

(4) We now continue the original computation with Figure 47.

To make our idea crystal clear, let us present the reasoning structure as follows: (Note the network changes as we reason, so each line has to indicate which network we are dealing with).

1. Figure 45, $a =$ in, assumption
2. Figure 45, $b =$ out, from (1)
3. subcomputation in Box1

Box 1:

3.1	Figure 45, $c =$ in, assumption
3.2	Figure 47, $c =$ out, from (3.1)
3.3	Use (3.1) and (3.2) and the Shkop principle: c must be out
3.4	Exit subcomputation with Figure 47

4. We are now in Figure 47 from Box 1: we continue reasoning.

   Thus the network changes as we reason along the arrows!

At present we do not know if this new computation is sound. It may be that counter examples can be found. Even if it is sound, we do not know exactly what it does. Our conjecture is that it just forces us to consider the loops first and eliminate them. At any rate, this is not crucial to our section, since we are happy with the General Shkop Principle and the algorithm we have in the Appendix.

5.3 Conclusion

The original Shkop principle is given in temporal context: Imagine a group of agents operating in time and taking actions. To execute an action **a** the pre-condition of the action $\alpha_{\mathbf{a}}$ needs to be fulfilled and then after the action is taken, the post-condition $\beta_{\mathbf{a}}$ of the action holds.

So if we start at time $t = 0$ in a certain state s and let our agents proceed with their actions then we move from state to state without any trouble and no argumentation networks arise and no loops arise.

The difference between ordinary actions and Talmudic actions is that the Talmud allows for the pre-conditions to contain future conditions and actions. Thus the enabling conditions of the actions can depend on the future and this can create loops. The Talmud also says that if the future condition is not fulfilled, then the action is nullified backwards in time (backward causality). As discussed in previous subsections.

To give a simple example, suppose that on Monday John orders a new laptop to be delivered on Friday. John gives his old laptop on Monday to a student named Tracy, free of charge, on the condition that on Friday, Tracy will configure his new laptop. Call this action **a**.

On Tuesday Tracy is ready to sell the laptop she got from John to a new buyer, Mary, for a good price, but Mary insists that on Friday Tracy transfers the contents of her old computer to the laptop she is buying. Call this action **b**.

The pre-condition for action **b** is that Tracy owns the laptop she is selling. For this to hold she must configure John's new laptop on Friday. However, if we allow action **b**, then Tracy will not be able to configure John's new laptop on Friday, because she will be busy transferring Mary's old data. If Tracy does that for Mary, then action **a** is nullified and so Tracy will not be the owner of the laptop she wants to sell and therefore action **b** is nullified.

What we get here is that if action **b** is allowed then it is nullified. The Shkop principle says that in this case do not allow action **b**.

We see here the context in which the Shkop principle operates. It is a time action model with future pre-conditions and backward causality, which progresses in time. Shkop says that any action which is about to be taken at time t which causes a chain reaction which cancels its own pre-condition at the same time t, should not be taken at time t.

We used the idea of Shkop to suggest and create the Shkop extensions for argumentation networks. These networks are not temporal but are static. We get them from the temporal action model by looking at what is happening at any certain fixed time.

This initial discussion is mainly qualitative and a more detailed modelling of the temporal aspects is forthcoming.

APPENDICES

APPENDIX A: TABLEAUX FOR CAMINADA SOCRATIC DISCUSSION

We offer here a tableaux method designed to test, for an element x in a finite argumentation network, whether x is an element of any admissible extension. Compare also with the Verheij paper [107].

DEFINITION 73. Let $\mathbf{N} = (S, R)$ be a finite argumentation frame.

1. A tableaux for $\mathbf{N}$ has the form

$$\tau = (\mathbb{A}_\tau, \mathbb{B}_\tau, \mathbb{D}_\tau)$$

   where $\mathbb{A}_\tau \subseteq S$ is the left inside of τ and $\mathbb{B}_\tau \subseteq S$ is the right outside of τ, and $\mathbb{D}_\tau$ is the set of elements marked to be treated in τ. $\mathbb{D}_\tau$ will be treated in the next tableau derived from τ. We have either $\mathbb{D}_\tau \subseteq \mathbb{A}_\tau$ (left treatment) or $\mathbb{D}_\tau \subseteq \mathbb{B}_\tau$ (right treatment).

2. A tableau τ is said to be closed if one or more of the following holds:
   - $\mathbb{A}_\tau \cap \mathbb{B}_\tau \neq \varnothing$
   - For some $y \in \mathbb{B}_\tau$, we have $\{x \in S | xRy\} = \varnothing$.

DEFINITION 74. Let $\mathbf{N} = (S, R)$ be finite argumentation frame and let $x \in S$. We define a tree $\mathbb{T}$ of tableaux for testing whether $x =$ in is possible at all, i.e. whether x can be a member of any admissible extension. The tree of tableaux will have tree relation ρ.

Step 1
Form the tableau $\tau_1 \in \mathbb{T}$, where

$$\tau_1 = (\{x\}, \varnothing, \{x\})$$

say $\{x\}$ is marked to be dealt with at this stage.

Step 2
Form the tableau $\tau_2 \in \mathbb{T}$, where

$$\tau_1 = (\{x\}, \{y | yRx\}, \{y | yRx\})$$

Say $\{y | yRx\}$ are marked to be dealt with at this stage and that $\{x\}$ has been dealt with. Let $\tau_1 \rho \tau_2$ hold.

If for some y such that yRx we have $\{z | zRy\} = \varnothing$ or if xRx then this tableau is closed. Otherwise we move to Step 3.

Step 3
Let $\mathbf{f}$ be any choice function such that for each y to be dealt with in the

tableaux τ_2 of the previous step, (i.e. $y \in \mathbb{D}_{\tau_2}$), it chooses an element $\mathbf{f}(y) \in S$ such that $\mathbf{f}(y)Ry$. Form the tableaux, $\tau_3^{\mathbf{f}} \in \mathbb{T}$:

$$\tau_3^{\mathbf{f}} = (\mathbb{A}_3^{\mathbf{f}}, \mathbb{B}_3^{\mathbf{f}}, \mathbb{D}_3^{\mathbf{f}})$$

for each such an $\mathbf{f}$, where

$$\begin{aligned}
&\mathbb{A}_3^{\mathbf{f}} = \mathbb{A}_2 \cup \{\mathbf{f}(y) | y \in \mathbb{B}_2\} \\
&\mathbb{B}_3^{\mathbf{f}} = \mathbb{B}_2 \\
&\mathbb{D}_3^{\mathbf{f}} = \{\mathbf{f}(y) | y \in \mathbb{B}_2 \text{ and } \mathbf{f}(y) \notin \mathbb{A}_2\}.
\end{aligned}$$

Say that all elements of $\mathbb{B}_2$ (all the ys) have been dealt with and all elements of $\mathbb{D}_3^{\mathbf{f}}$ are marked to be dealt with.

Let $\tau_2 \rho \tau_3^{\mathbf{f}}$, for all $\mathbf{f}$.

Note that $\mathbb{D}_3^{\mathbf{f}}$ may be empty.

Step 4
Let $\tau_3^{\mathbf{f}}$ be any tableau of Step 3. Construct the tableau $\tau_4^{\mathbf{f}} \in \mathbb{T}$ as follows:

$$\begin{aligned}
&\mathbb{A}_4^{\mathbf{f}} = \mathbb{A}_3^{\mathbf{f}} \\
&\mathbb{B}_4^{\mathbf{f}} = \mathbb{B}_3^{\mathbf{f}} \cup \{z | \text{ for some } u \in \mathbb{D}_3^{\mathbf{f}} \text{ we have } zRu\} \\
&\mathbb{D}_4^{\mathbf{f}} = \{z | \text{ for some } u \in \mathbb{D}_3^{\mathbf{f}} \text{ we have } zRu \text{ and } z \notin \mathbb{B}_3^{\mathbf{f}}\}.
\end{aligned}$$

We say the elements of $\mathbb{A}_3^{\mathbf{f}}$ have been dealt with and the elements of $\mathbb{B}_4^{\mathbf{f}}$ are marked to be dealt with.

Let $\tau_3^{\mathbf{f}} R \tau_4^{\mathbf{f}}$.

Inductive step type odd
We assume by induction that we have $\tau = (\mathbb{A}, \mathbb{B}, \mathbb{D})$ and the elements marked to be dealt with are all in $\mathbb{A}$, i.e. $\mathbb{D} \subseteq \mathbb{A}$ and $\mathbb{D} \neq \varnothing$. In this case proceed as in Step 3 and create τ' and let $\tau' \in \mathbb{T}$ and let $\tau R \tau'$.

Inductive step type even
We assume by induction that we have $\tau = (\mathbb{A}, \mathbb{B}, \mathbb{D})$ and all the elements to be dealt with are from $\mathbb{B}$ (i.e. $\mathbb{D} \subseteq \mathbb{B}$), and that $\mathbb{D} \neq \varnothing$.

Then proceed as in Step 4.

LEMMA 75. *If $\mathbf{N} = (S, R)$ is finite then after a finite number of steps the Tableaux process terminates. We reach tableaux at the bottom of the ρ-tree such that they are either closed or their $\mathbb{D}$ is empty.*

Proof. Since $\mathbb{D}$ always adds new elments either to $\mathbb{A}$ or to $\mathbb{B}$ and $\mathbb{A}$ and $\mathbb{B}$ do not decrease, and S is finite, sooner or later $\mathbb{D} = \varnothing$. ■

LEMMA 76. *Let (S, R) be a finite argumentation network and let $(\mathbb{T}, \rho)$ be the tableaux for it.*

Then there exists a maximal path $\tau_1 \rho \tau_2 \rho \ldots \rho \tau_n$ of non-closed tableaux in $\mathbb{T}$, if and only if x is a member of some admissible extension E.

Proof.

1. Assume $x \in E$ and E is an admissible extension. We will define a maximal path $\tau_1 \rho \tau_2 \rho \ldots \rho \tau_n$ of non-closed tableaux in $(\mathbb{T}, \rho)$.

   Let $\tau_1 = (\{x\}, \varnothing, \{x\})$ as in Step 1 of the inductive definition of $(\mathbb{T}, \rho)$.

   Let τ_2 be as in Step 2. τ_2 is not closed, because if xRx holds, then x cannot be in any admissible extension, and if for some y, yRx and $\neg\exists z(zRy)$ hold, then x is out.

   Assume by induction that we have defined a chain $\tau_1 \rho \tau_2 \rho \ldots \rho \tau_k$ of non-closed tableaux such that for each $1 \leq i \leq k$ we have

   - If $y \in \mathbb{A}_{\tau_i}$ then $y \in E$
   - If $y \in \mathbb{B}_{\tau_i}$ then for some $z \in E, zRy$ holds.

   We now define τ_{k+1}.

   Case k is odd
   In this case we have
   $$\mathbb{D}_{\tau_k} \subseteq \mathbb{A}_{\tau_k}$$
   Let τ_{k+1} be defined in Inductive Step type odd (same as Step 3). Clearly $\tau_k \rho \tau_{k+1}$ holds. We want to show that τ_{k+1} is not closed. Since $\mathbb{A}_{\tau_k} \subseteq E$ and $\mathbb{D}_{\tau_k} \subseteq \mathbb{A}_{\tau_k}$ we have that any yRu for $u \in \mathbb{D}_{\tau_k}$ is atatcked by E and hence is out. Thus
   $$\mathbb{A}_{\tau_{k+1}} \cap \mathbb{B}_{\tau_{k+1}} = \varnothing.$$
   Also every such y is attacked by something and so τ_{k+1} is not closed.

   Case k is even
   In this case we have $\mathbb{D}_{\tau_k} \subseteq \mathbb{B}_{\tau_k}$. This means that all points of $\mathbb{D}_{\tau_k}$ are out. Moreover by construction, $\mathbb{D}_{\tau_k}$ are points attacking points in $\mathbb{A}_{\tau_{k-1}}$, and so by the admissibility of E each such point y has an attacker $\mathbf{f}(y) \in E$. Then let τ_{k+1} be $\tau_{k+1}^{\mathbf{f}}$ for this function $\mathbf{f}$. we have that $\tau_k \rho \tau_{k+1}^{\mathbf{f}}$ and $\tau_{k+1}^{\mathbf{f}}$ is non-closed.

   We carry on until such an n that $\mathbb{D}_{\tau_n} = \varnothing$.

2. Assume there exists a maximal path of non-closed tableaux $\tau_1 \rho \tau_2 \rho \ldots \rho \tau_n$ in $(\mathbb{T}, \rho)$. Then clearly
   $$\mathbb{D}_{\tau_n} = \varnothing.$$
   Let $E = \mathbb{A}_{\tau_n}$. We show that E is conflict free and self-defending. If xRy holds for $x, y \in E$, then at some $\tau_i, y \in \mathbb{A}_{\tau_i}$ and so $x \in \mathbb{B}_{\tau_{i+1}}$ and so τ_j will be closed, for some $j \geq i$ (the j in which x gets into $\mathbb{A}_{\tau_j}$).

   Assume for some z that $zRx, x \in E$. We need to show a $u \in E$ such that uRz. Since $x \in E$ then $x \in \mathbb{A}_{\tau_i}$ for some i. Then $z \in \mathbb{B}_{\tau_{i+1}}$ and so in $\mathbb{B}_{\tau_{i+1}} = \mathbb{B}_{\tau_i}\mathbf{f}$ we have $\mathbf{f}(z) \in \mathbb{A}_{\tau_i} = \mathbb{A}_{\tau_{i+1}}$ and $\mathbf{f}(z)Rz$.

This completes the proof. ■

EXAMPLE 77. Let us check again whether $c =$ in is possible in Figure 31, this time using tableaux.

$$\begin{array}{l} \tau_1 : (\{c\}, \varnothing, \{c\}) \\ \tau_2 : (\{c\}, \{a, b\}, \{a, b\}) \\ \tau_3^{\mathbf{f}} : (\{c, a, b\}, \{a, b\}, \{a, b\}). \end{array}$$

Here $\mathbf{f}(a) = b$ and $\mathbf{f}(b) = a$. $\tau_3^{\mathbf{f}}$ is closed.

REMARK 78. Note that the tableaux method works for the query for several points, namely

- Can $c_1, \ldots, c_n$ all be together in some admissible set?

We simply start our tableaux with

Step 1:

$$(\{c_1, \ldots, c_n\}, \varnothing, \{c_1, \ldots, c_n\})$$

5.4 Appendix B: Shkop principle in temporal context

It would be helpful to the reader if we present the Shkop prinicple in its natural temporal context. Imagine a linear flow of time of the form $(N, <)$ where N is the set of natural numbers $\{1, 2, 3, \ldots\}$ and $<$ is smaller than relation. We associate with each $n \in N$, a state of the world, which we denote by Δ_n, being a classical propositional logical theory in the language with the atoms $Q = \{q_1, \ldots, q_k\}$. We imagine history as evolving. At step 1 we have only state Δ_1 as given and state Δ_2 has not been created yet. The future states are created by actions. An action has the form $\mathbf{a} = (\alpha_{\mathbf{a}}, \beta_{\mathbf{a}})$, where $\alpha_{\mathbf{a}}$ is the precondition of the action and $\beta_{\mathbf{a}}$ is the post condition, all in the same classical langauge of the states Δ.

So at state Δ_1 we might wish to take action $\mathbf{a}$. We can do that if the precondition holds, i.e. $\Delta_1 \vdash \alpha_{\mathbf{a}}$. If this is the case, then we take the action and we move to state Δ_2, which is the state at time 2. Δ_2 is connected with Δ_1 via a revision process, denoted by "$\circ$". Thus $\Delta_1 = \Delta_1 \circ \beta_{\mathbf{a}}$.

The exact nature of the revision process is not relevant to our purpose. It is sufficient to know that for any Δ and any action $\mathbf{a} = (\alpha_{\mathbf{a}}, \beta_{\mathbf{a}})$, such that $\Delta \vdash \alpha_{\mathbf{a}}$ we get a new state $\Delta' = \Delta \circ \beta_{\mathbf{a}}$.

This is a simple model which can easily be made richer and more complicated. The Talmudic twist to this model is that the Talmud allows for future *conditional actions*. Part of the precondition for allowing action $\mathbf{a}$ to take place at time n is that a related action $\mathbf{a}'$ be taken at future time $n + n'$.

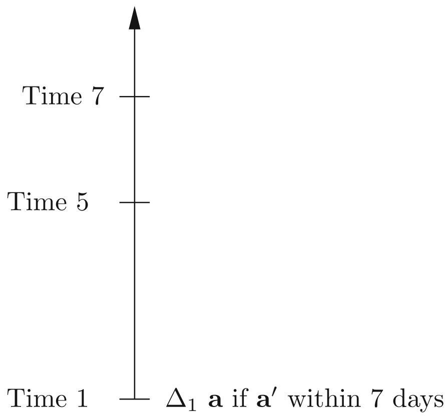


Figure 48.

For example: I give you this computer to be yours now on the condition that you clean my garden in a week's time. We have

$\mathbf{a}$ = (I own computer, you own computer)
$\mathbf{a}'$ = (Truth, you have cleaned my garden)

So if you do not clean my garden in a week, then the original action is cancelled. This is *backward causality*. We denote these conditional actions by **a** if a' within n' days.

Consider now Figure 48.

So action **a** can be taken at Δ_1 on the condition that action $\mathbf{a}'$ is taken at time 7. Note that time 7 has not yet happened.

Now suppose we continue to take actions and at time 5 we want to take action **b**. The precondition of action **b** holds and so we want to proceed. It could be the case that if we take action **b** at time 5, then a situation is created where action $\mathbf{a}'$ cannot be taken at time 7. If action $\mathbf{a}'$ is not taken at time 7, then action **a** at time 1 is not valid and past history is affected to the extent that at time 5 in the new history, the precondition for action **b** does not hold. The Shkop principle says that any action **b**, which when taken, changes history backwards in such a way that it cannot be taken (cancelling its own precondition) then **b** should not be taken!

A simple example will illustrate the idea:

EXAMPLE 79. On Monday, John buys a new computer to replace his old one. He gives the old computer (which is still good and fast) to his student Terry, on the condition that on Saturday, Terry comes to John's home and installs the new computer.

Terry decides to sell the computer he was given to a housewife neighbour called Mary. The precondition for the sale is that Terry owns the computer. This is OK because there is still the possibility for Terry to fulfil the condition to John and go on Saturday and install John's new computer.

Mary is prepared to buy the computer from Terry but she has her own condition. She wants Terry to come on Saturday and teach her how to use it.

We ask: can Terry sell the computer to Mary? We reason, following Shkop, that if Terry does sell the computer to Mary, he will have to spend Saturday with her and would not be able to go to John and install John's new computer. Failing to go to John on Saturday would nullify the gift of John giving Terry the old computer, which would nullify the precondition of the sale of this computer by Terry to Mary, namely Terry is not the owner of this computer.

So by selling the computer to Mary, Terry is nullifying the legitimacy of the sale!

So the Shkop principle applies and the sale cannot be permitted.

Let us now give another temporal argumentation model in which the Shkop principle can apply for resolving loops.

Suppose we have a sequence of argumentation networks of the form $(S_n, R_n), n = 1, 2, 3, \ldots$ such that $S_n \subseteq S_{n+1}$ and $R_n \subseteq R_{n+1}$. Thus as time passes on, (i.e. $n = 1, 2, \ldots$) we get more and more arguments and more and more attacks.

Consider time $n + 1$ and let $x \in S_{n+1} - S_n$. So x is a new argument added at time $n + 1$. So if x causes an odd loop and cannot be part of any extension, then we apply the Shkop principle, as detailed in Section 2, and annihilate it. To understand the usefulness of this principle and the temporal setup, consider Figure 38. We have many options for resolving the loops there. Our task is made easier if we have a temporal sequence of when each argument was put into the figure. We can follow the temporal sequence and use the Shkop principle to incrementally in time resolve the loops.

6 UNCERTAINTY RULES IN TALMUDIC REASONING

INTRODUCTION

The Babylonian Talmud, compiled from the 2nd to 7th centuries CE, is the primary source for all subsequent Jewish law. It is not written in apodictic style, but rather as a discursive record of (real or imagined) legal (and other) arguments crossing a wide range of technical topics. Thus, it is not a simple matter to infer general methodological principles underlying the Talmudic approach to legal reasoning. Nevertheless, in this section, we proposal a general principle that we believe helps explain the variety of methods used by the Rabbis of the Talmud for resolving uncertainty in matters of Jewish

Law (henceforth: halacha). Such uncertainty might arise either if the facts of a case are clear but the relevant law is debatable or if the facts themselves are unclear.

A Formal Model

Roughly speaking, the principle we argue for is that, in general, halachic rules for dealing with uncertainty are not probabilistic, but rather are action rules telling us what to do.

Thus, suppose that in situation S we have that

1. If a_1 we do x_1
2. If a_2 we do x_2
3. $\neg x_1 \wedge \neg x_2$ implies y

If there is a 50% doubt about a_i we formally decide $\neg x_i$. Assume that there is such a 50% doubt. Having formally decided $\neg x_1 \wedge \neg x_2$, we now get y. This conclusion holds even if we know that $a_1 \vee a_2$ must logically hold.

Let's now consider one such model.

Our starting point is a language for describing states and actions. Our language has constants for states, $s_1, \ldots, s_k, \ldots$, constants for actions $a_1, a_2, \ldots$ and notation for sets of actions, e.g. $\mathbb{A} = \{a_1, a_2\}$.

We can take predicates like $P(s, x)$ meaning $P(x)$ holds at state s, and predicates like $\mathbf{move}(\mathbb{A}, s, s')$ reading: the set of actions $\mathbb{A}$ moves us from state s to state s'. Our axioms have the form:

$$\bigwedge_i P_i(s, x) \wedge \mathbf{move}(\mathbb{A}, s, s')) \to \bigwedge_j P'_j(s', x).$$

If state s satisfies P_i for x and we move to s' by doing $\mathbb{A}$, then s' satisfies P'_j for x.

We also have a language with meta-predicates $\Psi(P(s, x))$ reading: the property P is classified as an instance of Ψ, at the state s for the individual x.

A history $\mathbf{s}$ is a sequence of states $\mathbf{s} = (s_1, s_2, \ldots, s_k)$ such that $\mathbf{move}(\{a_i^j\}, s_i, s_{i+1})$ holds for some $j = 1, 2, \ldots, m(i)$. This means the actions $\{a_i^j\}, j = 1, 2,$ were taken at s_i and we shifted from state s_i to state s_{i+1}. We also have a language with O_T and F_T. $O_T\Psi$ means Ψ is obligatory and $F_T\Psi$ means Ψ is forbidden. For the nature of rules and character of O and F see [5].[83]

A halachic decision takes the following form. Suppose we moved along the states $s_1, \ldots, s_r$. Suppose at state s_i we have $P_{i,j}(s_i, x_i)$ holding, $j =$

[83] Note that here we use the meta-predicates Ψ as objects of obligations and prohibitions $O_T\Psi$ and $F_T\Psi$. We can use predicates $P(s, x)$ as well as meta-predicate defining

$1, \ldots, m(i)$. Suppose actions $\{a_i^k\} = \mathbb{A}_i$, $k = 1, \ldots, n_i$, is responsible for moving us from state s_i to s_{i+1}. Then the Halacha might stipulate that $\Psi(P(s_r, x))$ holds. We write this as follows

(*)

$$\bigwedge_i \mathbf{move}(\mathbb{A}_i, s_i, s_{i+1}) \wedge \bigwedge_{i,j} [P_{i,j}(s_i, x_i) \text{ and } \Psi_{i,j}(P_{i,j}(s_i, x_i))] \Rightarrow! \Psi(P(s_r, x).$$

the arrow '⇒!" symbolises halachic stipulation and we allow for some of the $\Psi_{i,j}$ not to appear in (*).

Thus, for example, the Bible forbids doing any work on the Sabbath. Call this $F_T\Psi_1$ where $\Psi_1(P)$ means that P is a "work" predicate. A fellow bought a complicated do-it-yourself cupboard and wants to slot all pieces together on the Sabbath. Is this considered work? Here $P(x)$ is "to slot x together" and we are asking whether $\Psi_1(P(x))$ holds. Once a ruling is given, then the ruling holds from then on, and $P(x)$ is forbidden.

Let $\mathbf{HR}(s)$ be the set of Halachic rulings of the form (*) available at state s. When we move from state s to state s', we carry the Halachic rulings with us and may add some new rules. Thus $\mathbf{HR}(s)$ is a subset of $\mathbf{HR}(s')$. When in state s' a question arises as to the status of some predicate $\Psi(P(s, x))$, we check whether some ruling of the form (*) can be instantiated to give an answer. If not we have to ask the Rabbis for a ruling and the new ruling of the form (*) is added to $\mathbf{HR}(s')$. This is how the sets $\mathbf{HR}$ grow and evolve

Majority Rules in the Talmud

In what follows we treat various uncertainty examples in the Talmud and show that the considerations involved are not probabilistic but operational rulings for the practicing individual to take action.

One of the Talmud's guiding principles for dealing with uncertainty is "follow the majority" (Hullin 11a). As we shall see, this rule is applied in a variety of ways. Perhaps the canonical form of the rule concerns the oft-cited (e.g., Ketubot 15a) case in which an unlabelled piece of meat is found on the street in a town with p kosher butcher shops and q non-kosher butcher shops. All other considerations (such as proximity and size of the shops) being equal, the meat is deemed kosher if and only if $p > q$. We note as an aside that will be of some importance below that if $p = q$, the meat is deemed "in doubt" by this decision method and a secondary decision method must be invoked to resolve the matter.

---

themselves., i.e.

$$\Psi_{P(s,x)}(Q(s', y))$$

holds iff

$$Q(s', y) = P(s, x)$$

holds. Thus we can also have $O_T P$ and $F_T P$

This sounds rather straightforward. However, scattered around the Talmud we find a number of extensions of this rule as well as a number of limitations. Let's now consider the complete picture.

The Talmud (Tractate Hullin 11a) states that there are two distinct principles of "follow the majority" that cannot be inferred one from the other. The first principle is typified by the example of the meat we just considered in which the majority is said to be "present". The second principle involves what is called an "absent" majority and is typified by the following example. The milk of a cow suffering from some life-threatening illness is not kosher. Such an illness might be completely undetectable unless we slaughter the cow and perform an autopsy. Nevertheless, we can drink milk from a random cow, despite the inevitable uncertainty regarding its health, because most cows are healthy.

The Talmud does not define the difference between present and absent majorities but it is worth attempting to define that difference since there are important differences in the respective follow the majority principles. For example, the 2nd century scholar, Rabbi Meir, asserts that an absent majority does not constitute sufficient grounds to overturn an existing status quo but implies that a present majority does (Yevamot 67b). Conversely, in capital cases, where we require something approximating certainty to convict, a present majority never constitutes grounds for conviction but an absent majority might. For example, in a case that rests on establishing the identity of a defendant's mother, the fact that 99 out of 100 candidate mothers would satisfy the conditions for conviction (a present majority) is inadequate grounds to convict, but the fact that (in the absence of countervailing evidence) most apparent family relationships represent biological relationships (an absent majority) is sufficient grounds to convict.

The examples cited in the Talmud of each type of majority as well as the above rules suggests that the difference is that a present majority entails a closed set of objects the proportion of which have some relevant status is known. An absent majority entails some empirical claim regarding the proportion of a population that has some relevant status, where the claim is based on some sample. Since this sample might not be currently present, such a majority is regarded as absent. (The distinction between present and absent majorities can be fruitfully compared with the distinction between the classical interpretation of probability, motivated by gambling applications, and the frequentist interpretation of probability, motivated by insurance applications. See [87].)

Consequently, the rule that we follow a present majority is regarded as formal and procedural. It is treated identically whether the majority is 0.51 or 0.99. Furthermore, the conclusion to which it leads is never regarded as a certainty sufficient for convicting in a capital case. By contrast, an absent majority is tied to an underlying empirical claim and hence the rule that we follow an absent majority is linked to the strength that the Rabbis wished to assign to that claim. Rabbi Meir always regards empirical claims as sufficient only to yield a default rule, which in turn he regards as no stronger than a different default rule that presumes that the last known status quo continues. On the other hand, those who do not accept Rabbi Meir's view hold that a sufficiently strong empirical claim be treated as a certainty for legal purposes.

Let's now consider a simple paradox that arises in the use of the rule that we follow an absent majority.

1. Suppose that known kosher milk and known non-kosher milk (call this state s), have been inadvertently mixed. Call this action of mixing action a, resulting in a new state s'. In symbols, we have

   $\neg$ Kosher (s, unit of milk with label number i) $i = 1, 2, 3, \ldots, n$, and Kosher (s, unit of milk with label number j), $j = n + 1, n + 1, \ldots, m$ and action **Mix** applied to the units takes us from state s to the new mixed state s' where the numbering labels on the units is lost.

   We need a rule which will say whether Kosher (s', unit of milk without a number) is true or not in state s'.

2. The mixture is kosher if the proportion of kosher to non-kosher units milk in the mixture is greater than 60:1. This is the rule applied in his case which decides whether the milk is kosher. So the rule is

   $\bigwedge \neg$ Kosher (s, unit of milk with label number i) $i = 1, 2, 3, \ldots, n$, and $\bigwedge_j$ Kosher (s, unit of milk with label number j), $j = n+1, n+1, \ldots, m$, and $n/m < 1/61$ and action **Mix** applied to the units takes us from state s to the new mixed state s' where the numbering labels on the units is lost $\Rightarrow$ Kosher (s', units of milk without a number).

   Now suppose that

3. it is known in general that 5% of all cows are non-kosher due to various endemic illnesses, though these cannot be identified through external examination. Call this state t.

4. Now we take the combined milk of a huge herd of cows, as is common in the dairy industry. Call this action b, resulting in state t'.

We need a rule to tell us whether this milk kosher or not. The probability that less than 1/61 of this milk is non-kosher is vanishingly small, so that one might think that it is non-kosher by the rule used in 2 above. Nevertheless, the vast majority of commentators do not rule this way. The principle is that by the rule that we follow an absent majority, it has already been decided (as we saw above) that each individual cow is kosher. Once that decision has been made, the matter is settled. The mixture is regarded as consisting of 100% kosher milk. Formally, the rule is as follows:

If in state s [less than 50% of cows are unhealthy] **and** [milk from unhealthy cow is non-Kosher] **and** [we take action **Mix** of all milk from any single cow at state s and thus move by **Mix** action to state s' in which the milk is mixed], **then** at s' the milk is Kosher.

We add the rule above to $\mathbf{HR}(s')$. This rule is of the correct form (*). Let us write it more carefully: If in state s [less than 50% of cows are unhealthy]. Call this $P(s,\ \text{cows})$ and [milk from unhealthy cow is non-Kosher] i.e $\neg\Psi(P'(s,\ \text{milk}))$ (where "Kosher" = "Ψ") and [we take action **Mix** of all milk from any single cow at state s, call this **move** $(\mathbf{Mix}, s, s')$ and thus move by **Mix** action to state s' in which the milk is mixed], then at s' the milk is Kosher i. e call the mixture $Q(s'\ \text{milk})$ **then** $\Psi(Q(s'\ \text{milk}))$.

If we write the ruling only in symbols we get

$$P(s,\ \text{cows}) \bigwedge \neg\Psi(P'(s,\ \text{milk})) \bigwedge\ \mathbf{move}(\mathbf{Mix}, s, s') \Rightarrow ! \Psi(Q(s'\ \text{milk}))$$

Extensions and Limitations of the Present Majority Rule

In what follows we discuss extensions and limitations on the formal decision rule that we follow a present majority. We will see that it too is quite different than what we customarily think of as probabilistic reasoning.

Suppose we have three pieces of identical meat of which one unidentified piece is non-kosher (call this state x). Using the "follow the (present) majority" rule, the Talmud states (Gittin 54b) that each of the pieces is regarded as kosher. More remarkably, the 15th century commentator Rabbi Asher (in gloss 37 to Hullin, Chapter 7) interprets this to mean that we are permitted to eat all three pieces simultaneously. Indeed, he rules that if the three pieces are liquefied, the liquid mixture can be drunk even though it is know with certainty that 1/3 of the mixture is non-kosher, far in excess of 1/61. The principle is quite clear. Once some rule has been invoked (in this case, to treat each piece as kosher), the matter is settled and can be applied even after subsequent state transitions occur such that the prior decision leads to absurd conclusions (in this case, that all the pieces can be eaten).

It is worth noting that if there are two pieces of identical meat of which one unidentified piece is non-kosher, call this state y, the majority rule is obviously inapplicable. In such case, we have an "unresolved set" and some secondary decision method must be invoked to resolve the matter. However, the secondary method invoked in the case of an unresolved set state y is different than the secondary method invoked in the case we saw above in which an isolated piece of meat is found in a town with an equal number of kosher and non-kosher butcher shops. The Talmud (Kritut 17b) does not treat an item from an unresolved set in the same way as it treats an item that is in doubt.

Another example of the present majority rule, indeed its purported source according to the Talmud (Hullin 11a), is the rule that when there is disagreement among a panel of judges, the ruling is according to the majority. The critical point to note is that in this case there is no uncertainty at all regarding facts and hence interpreting the present majority rule as a probabilistic method for resolving uncertainty regarding facts is, ipso facto, too narrow. Rather, the rule must be interpreted as concerning the treatment of mixed sets and can be stated as follows:

> Given a set of objects the majority of which have the property P and the rest of which have the property not-P, we may, under certain circumstances, regard the set itself and/or any object in the set as having property P.

An important extension of the present majority rule applies to cases in which the set in question does not consist of objects but rather of abstract possibilities. For example, in a civil case involving a man who accuses his wife of infidelity (based solely on the uncontested fact that at the time of their marriage she was not a virgin), the Talmud (Ketubot 9a) argues on her behalf that a) it is not known if she was raped or had intercourse of her own volition and b) in either case, it is not known if the event occurred subsequent to her contracting marriage with her husband. Since the husband would prevail in the case only for one of the four possibilities in the Cartesian product, he loses the cases on grounds of the formal present majority rule.

Formally the problem is that we know that at the time s' of the marriage x was not a virgin; call this $Q(s', x)$. Call her previous state s and assume that in that state she was a virgin but it is not clear what action moved her from state s to state s'. It could have been rape (action a_1, it could have been consent, action a_2 and either case could have been before or after their engagement, $P(s, x)$ or $\neg P(s, x)$. It is clear the formal pattern is the following:

- $P_i(s, x) \wedge \mathbf{move}(a_i, s, s') \to Q(s', x)$, for $i = 1, \ldots, k + m$.

We also know that

- $P_i(s, x) \wedge \mathbf{move}(a_i, s, s') \Rightarrow ! \Psi(Q(s', x)$, if $i \leq k$

and

- $P_i(s, x) \wedge \mathbf{move}(a_i, s, s') \Rightarrow ! \neg \Psi(Q(s', x))$, if $k < i \leq k + m$.

We observe $Q(s', x)$ but we do not know which action a_i was taken. What is the ruling Ψ or not Ψ?

The commentators note that no claim has been made that the probabilities of her having been raped or of her having had intercourse subsequent to the contract, respectively, are precisely $\frac{1}{2}$. Rather the claim is that nothing is known about these probabilities at all (and indeed if the probabilities were known, different decision method would be invoked). Thus, the method is vulnerable to manipulation in a manner somewhat reminiscent of Bertrand's paradox. For example, we could artificially collapse the majority argument by restating the crucial issue as "infidelity or not infidelity" and, conversely, we could artificially strengthen the majority argument (to seven out of eight, rather than three out of four) by distinguishing between violent rape and statutory rape. The argument is thus seen to rest rather formally on assumptions about what categories are natural kinds (e.g., rape) and which are not (e.g., statutory rape).

Finally, we turn to a crucial limitation on the application of the present majority rule. Suppose we have a set of ten pieces of meat, nine of which are kosher and one of which is non-kosher and identifiable as such (say, by its position in the pile). Now we randomly choose a piece without paying attention to which one, and, having done so, it is no longer possible to determine if it was one of the kosher pieces or the non-kosher piece. In contrast with the canonical case of a piece of meat found in the street in which we assign a status according to the majority of the sample from which it is drawn, in this case the Talmud rules that the proportion of kosher and non-kosher pieces is irrelevant. Rather, the set is treated as an unresolved set, precisely as in the case above in which two pieces of meat, one kosher and one non-kosher, are mixed.

The principle is this. An isolated item such as one found on the street must be assigned some status applicable to an individual item, e.g., kosher or non-kosher, and hence the majority rule is invoked to resolve the matter. An unresolved set might, however, be assigned a third status appropriate to a set, namely, neither kosher nor non-kosher but, rather, mixed. An item that is taken from an unresolved mixed set simply inherits the mixed status of the set from which it is taken; it is treated like a chip off the old block. Plainly, from a probabilistic point of view, it is hard to distinguish the case of the found piece from the case of the selected piece (and although it is tempting to suggest psychological explanations, these do not hold water when the full range of examples is carefully examined).

Conclusion

We have seen that the Talmudic way of dealing with uncertainty is pragmatic and not probabilistic

1. Given a situation s arising from action x, resulting in situation s' where some uncertainty occurs, the Talmud makes a decision that sticks and allows life to continue.

2. Given a situation x which could have arisen from one of actions $a_1, \dots, a_n$, we may have uncertainty as to the nature of the situa-

tion depending on which action a_i gave rise to it. Again, we use a rule to make a decision.

The above rules are not probabilistic because in a sequence of actions, we draw conclusions that persist even in cases where judging the final state in isolation might have lead to very different conclusions.

7 DELEGATION, COUNT AS, AND SECURITY IN TALMUDIC LOGIC

Introduction

Delegation is a commonplace feature in our society. Individuals give power of attorney to their lawyers to perform certain actions for them (e.g. buy or sell property), institutions delegate to certain employees to sign for them (human resources send letters of appointment) and owners can grant access and administrative rights to other people in relation to their servers.

The logic behind such a system has been studied by several communities.

In philosophy this is known as "count as". X counts as Y in context C.

In law there are various rules for power of attorney.

In computer science one talks about access control and delegation.

This section examines the approach to delegation in Talmudic Logic.

The current approaches to delegation, mainly study three features

1. Dominance — if several primary sources delegate to secondary sources who carry on delegating then what is the dominance relationship among the chains of delegations

2. Revocation — if some sources revoke the delegation or some change their minds and reinstate, how does this propagate through the chains of delegations?

3. Resilience — if one source revokes delegation do we cancel other delegations from other sources on the grounds that we now do not trust the delegate?

In the literature systems have been constructed which either model or implement a calculus of Delegation-Revocation (Privilege calculus). Their purpose is to answer the question of whether the chain of delegation and revocations can allow an agent to perform an action and their models are chain update models.

The Talmudic approach is slightly different not only in the details of its model but also in its point view.

The Talmud not only examines[84] the procedure of the actual acts of delegation and revocation and its calculus but also includes the analysis of ordinary actions (not just chain update actions) — their elements of agency, action, deliberation and competence. These attributes have preconditions addressing not only acts but also delegation and revocation chains leading to the actions. The Talmud also addresses cases of delegated agents unable to execute the actions for various reasons, and the possibility of agents going mad or dying during the delegation revocation process, with their repercussions.

7.1 Background and orientation

Delegation is a commonplace feature in our society. Individuals give power of attorney to their lawyers to perform certain actions for them (e.g. buy or sell property), institutions delegate to certain employees to sign for them (human resources send letters of appointment) and owners can grant access and adminisrtative rights to other people in relation to their servers.

The logic behind such a system has been studied by several communities.

In philosophy this is known as "count as". X counts as Y in context C [101; 102; 82], and see [72] for a survey.

In law there are various rules for power of attorney.

In computer science one talks about access control and delegation, see for example [97].

This paper examines the approach to delegation in Talmudic Logic.

The following are feature to be addressed:

1. The general logical context in which delegation takes place.
2. Exactly how (by what process) does agent **a** delegate to agent **b** item φ.
3. What are the rules for making chains of delegation?
4. How can delegation be revoked in a chain?
5. What happens in a delegation chain if some of the agents in the chain become insane (i.e. irresponsible or generally break down) and how to continue if such agents become sane again? What if they die (drop out permanently)?
6. What to do if some agents exceed their remit in a chain (e.g. human resources in a University sends a letter of appointment by mistake to the wrong candidate)?

[84] The Talmud deals with delegation — Shlichut — in various contexts. It appears in many places across Talmudic literature, for example Tractate Kidushin, 41b-42a (basic source on the subject), Tractate Gittin 18a, 33c (on inexplicit multiple delegation).

7. It may be the case that several agents $\mathbf{a}_i$ capable of executing action α, each delegates to the same agent $\mathbf{b}$ to do α. Meanwhile some of these agents $\mathbf{a}_i$ go insane, some die, and some cancel the delegation. What can $\mathbf{b}$ do?

A word on methodology. The Talmud (completed at the end of the fifth century) and its later interpreters (another 5-10 centuries) is full of debate about various cases of delegation. There is no formal logic, but a dialogue-based argumentation and analysis that is true to the casuistic nature[85] of the Talmud's core object of analysis (the Mishna). This process involves various case studies of "hard-cases" and approaches offered by various deliberators. Finding the logic behind such extravagantly lively debate, spread over thousands of discussions is challenging: It requires find a logical model with some degrees of freedom and a mapping of the various scholars or views to parameters in the logical model. We then have to go to all places and cases in the Talmud where there is a debate and the model must explain each move in each argument in each debate in each place in a perfect match. This is possible to do because as a body of law, Talmudic debates are remarkably coherent and consistent, with much effort invested in sorting out conceptual irregularities and disagreements. A formal-logical background is called for especially where it can benefit the Talmudic scholar (in the deliberation process) as well as the benefit of formulating that the Talmud implicitly uses and can be beneficial in the development of modern logic.The Talmudic logic project is geared toward this dual goal.

The topic of delegation is the sixth topic in Talmudic logic which we are examining. We have already published five books modelling five previous topics, using the same methodology.[86]

---

[85] Cf. Leib Moscovitz, Talmudic Reasoning: From Casuistics to Conceptualization, Tübingen: Mohr Siebeck, 2002.

[86]

1. Non-deductive Inference in the Talmud (with M. Abraham and U. Schild). 350pp, College Publications, 2010.
   We analyse the three basic non-deductive rules of Talmudic inference; namely Kal Vachomer (Argumentum A Fortiori) and the two kinds of Binyan Av (Analogy and Induction). We construct a unified Matrix Abduction model that explains all the major instances of these rules in the Talmud.

2. The Textual Inference Rules Klal uPrat. How the Talmud Defines Sets (with M. Abraham, G. Hazut, Y. Maruvka and U. Schild). 300pp, College Publications, 2010.
   We analyse the Klal uPrat family of textual rules in the Talmud. We view them as common-sense practical rules for defining sets. Such methods do not exist in general common-sense logical systems, and they complement the existing common-sense (non-monotonic) deductive logics.

3. Talmudic Deontic Logic (with M. Abraham and U. Schild). 296pp, College Publications, 2010.
   In this book we study the Deontic Logic of the Talmud. We find the system is different from the formal deontic logical system currently used in the general sci-

7.2 Motivating the Talmudic system

Let $\mathcal{A}$ be a set of actions and $\mathbf{A}$ be a set of agents. We need a relation $\mathbb{R} \subseteq \mathbf{A} \times \mathcal{A}$ giving us for each agent $\mathbf{a}$ in $\mathbf{A}$ the set of all actions $\alpha \in \mathcal{A}$ such that $\mathbf{a}$ can execute α ($\mathbf{a}$ has the authority to execute α). The actions have the form $\alpha = (A_\alpha, B_\alpha)$, where A_α is the pre-conditon and B_α is the post-condition. A_α and B_α are written in some predicate language $\mathbb{L}$, to be decided according to the required strengths and specifications of Talmudic delegation structure.

An agent $\mathbf{a}$ can delegate his authority to do any action to agent $\mathbf{b}$ (there are some restrictions on agent $\mathbf{b}$, like he has to be sane and responsible and can perform actions similar to α). He must not be involved in the action α himself, and the action α must be legally meaningful). We need a relation $\mathbb{D}$ where $\mathbb{D}(\mathbf{a}, \mathbf{b}, \alpha)$ means that $\mathbf{a}$ delegated action α to $\mathbf{b}$. This can be delegated further by $\mathbf{b}$. So the relation $\mathbb{R}$ can be expanded to a relation $\mathbb{F}^*$,

---

entific community, both in its ethical aspects as well as in its legal aspects. We show that the Talmudic distinctions between Obligations and Prohibitions are not based on the manner of execution of actions (positive action or lack of action) and offer a suitable model for such distinctions.Our model distinguishes between the normative and practical aspects of the Talmudic legal and ethical argumentation and discusses several applications and clarifications to current so called paradoxes of Deontic Logic as related to Contrary to Duties and to legal and ethical practical decision making.

4. Temporal Logic in the Talmud (with M. Abraham, I. Belfer and U. Schild). 674pp, College Publications 2011.
   This book studies Talmudic temporal logic and compares it with the logic of time in contemporary law. Following a general introduction about the logical handling of time, the book examines several key Talmudic debates involving time. The book finds that we need multi-dimensional temporal models with backward causation and parallel histories.

   It seems that two major issues are involved:

   (a) Actions conditional about future actions (Tenayim), connecting with backward causality;

   (b) Actions involving entities defined using future events (Breira), connecting with ideas from quantum mechanics.The book concludes with a general comparative discussion of the handling of time in general law and in the Talmud

5. Resolution of Conflicts and Normative Loops in the Talmud (with M. Abraham and U. Schild). 316pp, College Publications 2011.
   In this book we describe the fundamental rules for conflict resolution and address the basic Talmudic methods for resolving conflicts. We also investigate logical loops in Talmudic argumentation. It is obvious that one needs meta-level (out of the box) considerations. We also consider conflicts between Biblical Obligations and Prohibitions, a topic we studied in our third book. We conclude by comparing some features of conflict resolution with our matrix model presented in our first book.

namely

$$x\mathbb{R}^*\alpha \text{ iff } \exists y_1, \ldots, y_k \text{ for some } k, \text{ such that } y_1\mathbb{R}\alpha \wedge \bigwedge_{i=1}^{k-1} \mathbb{D}(y_i, y_{i+1}, \alpha) \wedge y_k = x.$$

In order to model the complexities of delegation in Talmudic logic we want to realise $\mathbb{D}$ using tokens (modern papers call them certificates, see for example [16; 43] and [96]).

An agent $\mathbf{a}$ which can do α has a token $\mathbb{T}(\mathbf{a}, \alpha)$. Think of it as a copy print of $(\mathbf{a}, \alpha) \in \mathbb{R}$. If $\mathbf{a}$ wants to delegate to $\mathbf{b}$, he signs on the token, "I authorise $\mathbf{b}$'. We denote this by $(\mathbf{a}, \mathbf{b}, \alpha)$. Thus we can get the chain $(y_1, \ldots, y_k, \alpha)$. If y_k wants to execute an action α, $\alpha = (A_\alpha, B_\alpha)$, the language $\mathbb{L}$ must also enable A_α to ask y_k: do you have a token $(y_1, \ldots, y_k, \alpha)$?

So for example to sell a table t, we need the agent $\mathbf{a}$ to own t and then he can sell it to agent $\mathbf{b}$. Or we can have $(\mathbf{a}, y_1, \ldots, y_k,$ sell table$)$ and y_k can sell the table on behalf of $\mathbf{a}$. So the language $\mathbb{L}$ must contain $\mathbb{D}$, as well as the names of agents and facts about the world.

This is a language where $\alpha = (A_\alpha, B_\alpha)$ and A_α can talk about α. It is a self reflecting language.

Different delegation theories will be implemented by different properties of the tokens.

There are two main types of delegation in Talmudic logic.

1. *Power of attorney* view (Maimonides[87] view).
2. The *long arm/extended reach* view (Tur[88] view).

If $\mathbf{a}$ delegates to $\mathbf{b}$ and $\mathbf{b}$ delegates further to $\mathbf{c}$, let us refer to $\mathbf{a}$ as the master (or principal, using modern terminology) and to $\mathbf{b}$ as the agent and to $\mathbf{c}$ as the subagent.

The *power of attorney* view is for the master to delegate to an agent to do action α.

The action is done by the agent and the result of the action is passed on to the master.

[87] Named here for Moses ben-Maimon, called Maimonides or Ramban (Hebrew acronym for "Rabbi Moshe ben Maimon"). For the *power of attorney* view, Cf. Rabbi Isaac Herzog, *The Main Institutions of Jewish Law, Vol. II. The Law of Obligations*, (2nd ed.) London: Soncino, 1967, pp. 141–142

[88] Jacob ben Asher, also known as Ba'al ha-Turim as well as Rabbi Yaakov ben Raash (Rabbeinu Asher), was likely born in Cologne, Germany, c. 1269 and likely died in Toledo, Spain, c. 1343. In point of fact, the *long arm/extended reach* and *power of attorney* concepts were suggested later on and used to explain the Tur. Cf. Ketzot Hachoshen (Aryeh Leib Heller-Kahane, 1745–1812), Ch. 188, ii; 244, iii.; Lekach Tov (Yosef Engel, 1858–1920), Ch. 1. In the Tur, the limits of delegation are mentioned in Even haEzer, Ch. 141 section 43, regarding the laws of divorce contracts (Gitin) and the ability of a receiving agent (Shaliach Kabala) to delegate his task.

Note by the way the English expression "The long arm of the law".

Owner of token: John Smith
Social Security number (SSN):

Action	Master (who has authority over the action)	Verify nomination of owner of this token (i.e. John Smith)
1. Sell house (property ♯)	Terry Jordan owner of the house	+from Terry
2. α	**a**	+from **a**

Figure 49.

action α. Issued by agent **a** such that $\mathbf{a}\mathbb{R}\alpha$.

Figure 50.

The *long arm* view is that the agent is an extension of the arm of the master, and the master is doing the action α by means of his arm extension — the agent.

So the agent "counts as" the master.

We model the difference between these two views through the properties of the token. The token is given from the master to the delegated agent and in the token there is a list of actions to be done by the agent.

The *power of attorney* view postulates a token for each agent. The *long arm* view postulates a token for each delegated action/job.

The token per agent view envisages the token as listing all the actions to be done. These include actions that the master has authority to do, as well as actions which he (the master) was originally recruited by a previous master to do (to whom he acts as an agent).

Figure 49 shows what this token looks like. Note also that this token allows for the master to cancel the appointment of the agent as an agent for the action (in modern terminology, the master revokes the delegation to the agent).

In the case of the *long arm* view, the tokens look like Figure 50

We now describe what happens when agent John Smith wants to execute an action. We check the following:

1. The *power of attorney* view checks whether action α shows in John Smith token (see Figure 49). Is he the master for this action? Was he appointed agent for this action by a master who has authority? Was he appointed by an agent who was himself appointed by a master?

etc. All the above is supposed to be recorded in the token.

2. The *long arm* view would simply check if our John Smith has the token as in Figure 50.

EXAMPLE 80. To see the difference between the two views, let us assume that the master **a** appointed **b** as an agent for him to do action α, and then lost his mind.

The *long arm* view will say the action cannot be performed because the source of the long arm, the master, is mentally incapacitated, making his *long arm/extended reach* useless, as he is without a sound mind. If we look at the token 50, the **a** in $\mathbf{a}\mathbb{R}\alpha$ is no longer sane.

The *power of attorney* view, the owner of the token (the agent **b**) is capable and sane, his token indicates he has the authority to take action, so he can do it!

Let us take a very simple example from practice. The manager of a company delegates to a secretary to delete certain sensitive files from the server, just before a shareholders' meeting is about to take place. The secretary goes to the meeting and intended to do the action afterwards. During the stormy meeting, the manager resigned and discussions were ongoing about appointing a new manager. The *long arm* view would say the secretary cannot delete the files because she is the long arm of the manager who is no longer in power, he resigned. The *power of attorney* view says that the secretary has a power of attorney, he/she should do the action and delete the files.

EXAMPLE 81 (Cancellation and reinstatement).

1. Both views allow for cancellation (the modern term is revocation). The master cancels either the token of Figure 49 or of Figure 50, depending on the view.

2. Both views agree that if the master becomes sane again (the manager of Example 80 gets reinstated) the action can take place without the need for doing again the formal appointment of delegation (i.e. the secretary need not ask the reinstated manager to reconfirm his instructions to him/her).

EXAMPLE 82 (The delegated agent becomes insane). Suppose the agent goes crazy, and then becomes sane again (goes through a mental breakdown for a while). Can he/she continue being an agent and execute the action?

According to the *long arm* view he can. He got the token, he is now sane, so he can do it. Similarly according to the*power of attorney* view. He got the token.

There is a difference however in the view about pre-condition A_α of α.

Owner of token: John Smith. SSN:

Action	Master **a**	who nominated John	confirm nomination from Levy not retracted	Who nominated Levy	Confirm nomination from Terry not retracted
sell house: address	Terry Jordan	+Levy =**c**	+confirm	+Terry =**b**	+confirm
α	**a**	**c**	+	**b**	(negative, **a** retracted the nomination

Figure 51.

The *long arm* view says A_α must check the sanity (capability) of both the master who controls the long arm and the agent, who is the arm. Both have to be functional.

The *power of attorney* view needs the functionality check in A_α of the agent only. The agent carries the token, he is supposed to execute the action!

We now examine how, according to each view, an agent can nominate a subagent for himself.

The *long arm* view treats this very simply. The agent has a token as in Figure 50. So he just passes this token on to his subagent. All very simple. The subagent is now the long arm of the master. The agent is no longer in the picture. We may have a long chain of such nominations. So a by-product of this view is that the master can cancel the nomination of his long arm agent at the end of the chain, no matter how long the chain is.

The *power of attorney* view would have to say that the subagent is delegated from the agent and not from the master. Thus the token must record this information. It must record the chain of delegations from agent to agent. A by-product of this view is that the master cannot cancel the nomination of the subagent. The subagent was nominated by the agent not by the master! The master can cancel the nomination of the agent but if the agent has already nominated a subagent then the nomination of the subagent stands! Figure 51 shows what the token looks like.

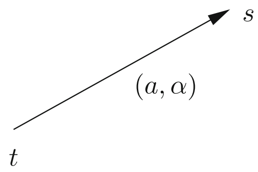


Figure 52.

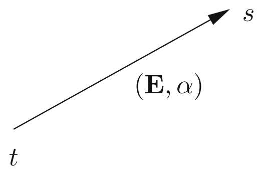


Figure 53.

7.3 Technical definitions of the logical model

7.4 Preliminary discussion

The framework in which we are working involves agents and actions, with a relation $\mathbb{R}$ between agents $\mathbf{a}$ and actions α, saying whether agent $\mathbf{a}$ can execute action α ($a\mathbb{R}\alpha$). α has preconditions A_α and postconditions B_α.

To this model we add the delegation component, in which agent $\mathbf{a}$ can delegate to agent $\mathbf{b}$ the execution of action α. We wrote this as $\mathbb{D}(\mathbf{a}, \mathbf{b}, \alpha)$.

Thus any model of delegation needs to be based on a model for agents and actions. Since our primary interest is in modelling delegation we can take a simple basic model of agents and actions, without any fine refinements (coming from a sophisticated multi-agent theories), provided that this agent-action model is rich enough to allow us to express all the delegation features we need to model.

DEFINITION 83. Let $\mathbf{A}$ be a finite set of agents and $\mathcal{A}$ a finite set of actions. By a basic multi-agent system we mean a tuple of the form $\mathfrak{M} = (S, \mathbf{R})$, where S is a non-empty set of states and $\mathbf{R} \subseteq (S \times S) \times \mathbf{A} \times \mathcal{A}$.

When $(t, s, \mathbf{a}, \alpha) \in \mathbf{R}$, we draw it graphically as in Figure 52. The figure means that at state t agent $\mathbf{a}$ can execute action α which moves the system to state s.

It may be that agent $\mathbf{b}$ can also execute α at state t but not agent $\mathbf{c}$. So we write Figure 53.

where $\mathbf{E} \subseteq \mathbf{A}$ is the set of agents which can execute α at state t.

Thus

$$\mathbf{E} = \{\mathbf{x} | (t, s, \mathbf{x}, \alpha) \in \mathbf{R}\}.$$

We need to require that the execution of α is deterministic, i.e.

- $(t, s, \mathbf{x}, \alpha) \in \mathbf{R}$ and $(t, s', \mathbf{y}, \alpha) \in \mathbf{R}$ impies $s = s'$.

We believe that this simple model is good enough for our purposes.

REMARK 84. Note that we ignored the representation of the preconditions A_α and postconditions B_α of actions α. This we can do because we use a set of states. If A_α does not hold at state t then the action cannot be taken. If it is taken then B_α holds at state s.

Thus to complete the model, we take $(S, \mathbf{R})$ and associate with each $t \in S$, a model $\mathbf{m}_t$ of the language $\mathbb{L}$ of the preconditions and postconditions, so that we can write $t \vDash A_\alpha$ or $s \nvDash B_\alpha$, etc.

Thus our final models have the form

$$\mathfrak{M} = (S, \mathbf{R}, \mathbf{m}_t), t \in S.$$

The above does not deal with delegation yet. We now examine our options for modelling delegation.

It is convenient to list the views and types of delegation we encountered in Section 2.

Type 1. **a** delegates to **b** action α in any context (state), e.g. "sell my house".

Type 2. **a** delegates to **b** action α only in certain contexts, e.g. "sell my house when the Euro is over 1.30 to the dollar".

Type 3. Preconditions and postconditions of actions involve delegation considerations, even though the action itself is not a delegation action.

For example, when **a** delegates to **b** action α then when **b** wants to execute α, part of the precondition of α is that agent **a** is sane and alive.

View 1. Agent **b** is the long arm of agent **a**.

In this case if we are at state t, we move to state s after the execution of the action (either by **a** or by **b**).

View 2. Agent **a** gives Talmudic power of attorney to agent **b** for example, to buy a house for him from agent **c** or to collect a debt for him from agent **c**.

In this case the execution of the action may result in an intermediate state. In the case of buying a house there is no intermediate state, but in the case of collecting a debt, where agent **c** also owes money to agent **b**, there is an intermediate state. We view the sequence of actions to be that the money first goes in the hands of agent **b**, who then passes it to agent **a**. So if we start at state t we move to t' and then to s, unlike the long arm delegation, where we move from t to s directly.

Option 1: The fibred (combined) option

This option puts a delegation program or logic next to a model $\mathfrak{M}$ for agents and actions. The program is used to update the relation $\mathbf{R}$ in $\mathfrak{M}$.

Let Δ be a database of all the delegation tokens in the system. The model becomes

$$\mathfrak{M} = (S, \mathbf{R}^{\Delta}, \mathbf{m}_t), t \in S.$$

When an agent delegates an action to another agent, the delegation database Δ is updated to Δ' and $\mathfrak{M}$ changes to

$$\mathfrak{M}' = (S, \mathbf{R}^{\Delta'}, \mathbf{m}_t), t \in S.$$

So the update system of Δ communicates with $\mathbf{R}$ of $\mathfrak{M}$.

Thus logically modelling such a system requires the logical modelling of communication between a program and a logic.

Since the updating system is independent of $\mathfrak{M}$, many researchers use Δ only as models of delegation and do not mention $\mathfrak{M}$ at all. This may not be possible if the systems interact. For example, part of the precondition of action α may be that it is not performed through delegation. The prime minister of a country or a King, for example, cannot freely delegate some actions associated with his position. A wife having difficulties giving a child to her husband cannot, nowadays, delegate the job to her maid, as was the custom in Biblical times.

Another example is when the delegations comes as a result of an action α in the real world is the following. If I run over the parents of a small child then by law the court becomes delegates for his interests. So here the delegation is a postcondition of my actions.

In such cases, where there is interaction between the delegation system and the preconditions and postconditions of ordinary actions, the next integrated model is a better option.

Option 2: The integrated model

This model views the delegation details (certificates, tokens, etc.) as part of the state $t \in S$, and the act of delegation is viewed as just another action modifying the state. So in this integrated model, $\mathbf{m}_t$ talks not only about facts, but includes the details of each delegation certificate/token ever issued. The precondition of actions must include that the action is executed by an agent who has the valid delegation token for the action.

A variety of models can be constructed under this option, but they all have the drawback that the delegation part gets a bit lost as a separate

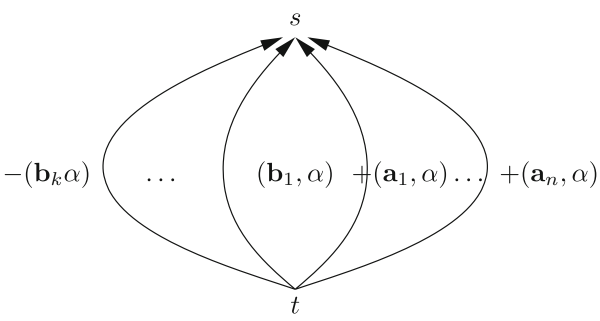


Figure 54.

system. We could save the uniqueness of the delegation part by separating delegation actions from ordinary actions and add temporal like connectives to the model which can chain correctly the delegation actions, etc., etc. If we do that well, we will be back to Option 1, hidden inside Option 2 under the guise of additional connectives.

Option 3: Reactivity model for the long arm *view of the Tur*

We now intuitively explain the components of this model. It is an integrated model which uses reactivity to separate delegation actions from ordinary actions in a natural way. For reactive Kripke models see [68].

It is best suited when there is interaction between delegation and post and preconditions of actions or where there are delegations which are valid only in certain contexts (states).

If delegations involve just agents and actions then Option 1 may be best.

First we decide that we keep the agent action model as it is, with the states describing pure facts, with no mention of delegation. So the relation $\mathbf{R}$ can be represented as in Figure 53. So there is no mention of delegation in this figure. We are going to add delegation into it. We need to write Figure 53 more explicitly. Consider Figure 54.

This Figure contains some information as Figure 53. We just wrote the information explicitly. We have $\mathbf{E} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\mathbf{A} - \mathbf{E} = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$.

The action α can be executed by $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and so we have the arrows $t \to s$ annotated by $+(\mathbf{a}_i, \alpha), i = 1, \ldots, n$. The action α cannot be taken by $\mathbf{b}_1, \ldots, \mathbf{b}_k$ and so we have the arrows $t \to s$ annotated by $-(\mathbf{b}_j, \alpha)$.

It would be easier to replace the relation $\mathbf{R}$ by its characteristic function $\mathbf{F_R}$. We have

$$\mathbf{F_R}(t, s, \mathbf{a}, \alpha) = 1 \text{ iff } (t, s, \mathbf{a}, \alpha) \in \mathbf{R}.$$

From now on we regard $\mathbf{R}$ as such a function (by abuse of notation).

We begin discussing delegation for the *long arm* view of the Tur:

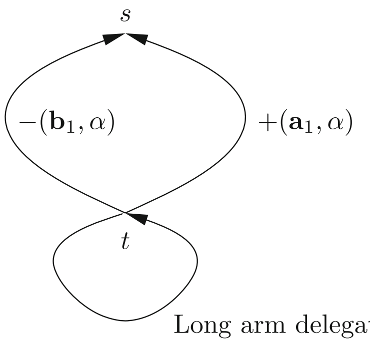


Figure 55.

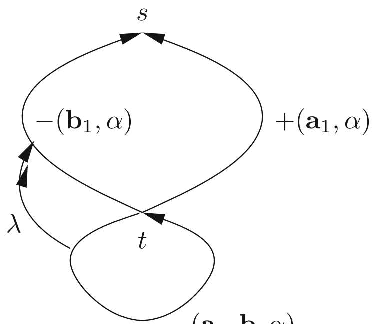


Figure 56.

Now suppose we perform an action of delegation. Agent $\mathbf{a}_1$, who can perform action α, wants to delegate to agent $\mathbf{b}_1$, the execution of α at state t. Agent $\mathbf{b}_1$ cannot perform action α at state t before the delegation (i.e. we have $+(\mathbf{a}_1, \alpha)$ and $-(\mathbf{b}_1, \alpha)$ before the delegation) but after delegation $-(\mathbf{b}_1, \alpha)$ is updated and changed to $+(\mathbf{b}_1, \alpha)$.

How do we represent this using reactive arrows? When we perform the delegation action, we remain at state t, since in our model, the states represent facts about the world and do not contain any token/certificate delegation information.

Figure 55 represents this move for $\mathbf{a}_1$ and $\mathbf{b}_1$.

We represent Figure 55 by the reactive Figure 56.

The reactive double arrow is written as

$$(t \to_{(\mathbf{a}_1, \mathbf{b}_1, \alpha)} t) \twoheadrightarrow_\lambda (t \to_{(\pm(\mathbf{b}_1, \alpha)} s).$$

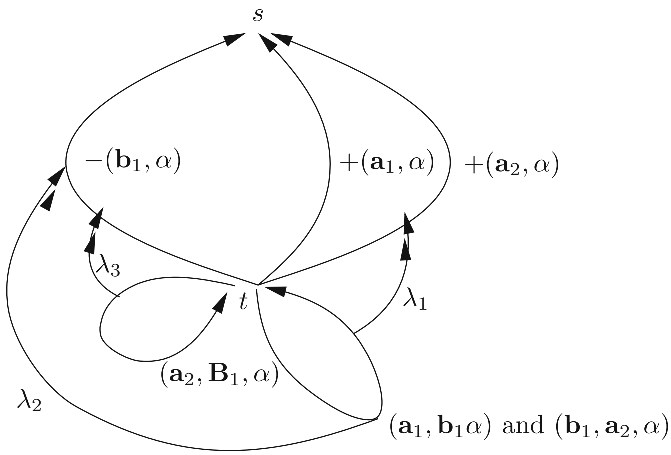


Figure 57.

λ is a label indicating the nature of the delegation/revocation double arrow. As $\mathbf{a}_1$ moves from t to t along the arrow, he triggers the double arrow which sends a signal λ to $t \rightarrow_{\pm(\mathbf{b}_1,\alpha)} s$. Let us assume that $\lambda =$ switch. Then if the annoation of $t \rightarrow s$ is "+", it turns into "-" and if it is "-" it turns it into "+". thus the double arrow is a switch!

Let us look at Figure 57.

Again, let us assume that $\lambda_1 = \lambda_2 = \lambda_3 =$ switch. Suppose $\mathbf{a}_1$ goes along the path

$$t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t$$

and then $\mathbf{a}_2$ continues along the path $t \rightarrow_{(\mathbf{a}_2,\mathbf{b}_1,\alpha)} t$.

The total movement is

$$t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t \rightarrow_{(\mathbf{a}_2,\mathbf{b}_1,\alpha)} t$$

The first $t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t$ switches $-(\mathbf{b}_1, \alpha)$ into $+(\mathbf{b}_1, \alpha)$. We consider that legitimate because we do have in the Figure $t \rightarrow_{+(\mathbf{a}_1,\alpha)} s$.

This same first movement also revokes the right of $\mathbf{a}_2$ to execute α. So it switches $+(\mathbf{a}_2, \alpha)$ into $-(\mathbf{a}_2, \alpha)$. Now $\mathbf{a}_2$ cannot delegate α because his ability to execute α was revoked by $\mathbf{a}_1$. Fortunately, $\mathbf{a}_1$ went again through $t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t$ and switched back to $+(\mathbf{a}_2, \alpha)$ and $-(\mathbf{b}_1, \alpha)$. Now $\mathbf{a}_2$ can go through his arc $t \rightarrow_{(\mathbf{a}_2,\mathbf{b}_1,\alpha)} t$ and switch on $+(\mathbf{b}_1, \alpha)$.

In the general case, where $\lambda_1, \lambda_2, \lambda_3$ can be general labels, not necessarily "switches", we need to collect the labels and decide whether the target arc (which is hit by several labels) is supposed to be on ("+") or not ("-").

For example, assume that

λ_1 = switch
λ_2 = dominant delegation
λ_3 = switch

So as we move along

$$t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t$$

the arc $t \rightarrow_{+(\mathbf{a}_2,\alpha)} s$ is switched to $t \rightarrow_{-(\mathbf{a}_2,\alpha)} s$ and the arc $t \rightarrow_{-(\mathbf{b}_1,\alpha)} s$ is changed to dominant $t \rightarrow_{+(\mathbf{b}_1,\alpha)} s$.

As we continue along

$$t \rightarrow_{(\mathbf{a}_1,\mathbf{b}_1,\alpha)\ and\ (\mathbf{a}_1,\mathbf{a}_2,\alpha)} t$$

the arc $t \rightarrow_{-(\mathbf{a}_2,\alpha)} s$ changes back to become $t \rightarrow_{+(\mathbf{a}_2,\alpha)} s$ as it is hit by λ_1 but the arc $t \rightarrow_{+(\mathbf{b}_1,\alpha)} s$ does not change as it is hit again by λ_2 = dominant delegation.

Now we continue along the arc

$$t \rightarrow_{(\mathbf{a}_2,\mathbf{b}_1,\alpha)} t$$

and the arc $t \rightarrow_{+(\mathbf{a}_1,\alpha)} s$ is hit by λ_3 = switch. the $+(\mathbf{b}_1,\alpha)$ does not change because it was hit before by λ_2 = dominant delegation and it is now hit by just a switch, which is not dominant.

Now if λ_3 were

λ'_3 = dominant revocation

then we would need to decide whether the triple $\{\lambda_2, \lambda_2, \lambda'_3\}$ should end up with + or with -.

This means that in the general case, when we go along a path and trigger various double arrows with labels, we need to calculate using a flattening algorithm Λ, whether any given arc is "+" or "-". We collect all the labels λ_j which hit the arc along the path and let Λ "flatten" it to either "+" or "-".

We also note, see Figure 58, that we have the notation to delegate from state t to state r, where r can be anywhere in the system. However the Talmud does not allow for delegation for action which is not definite now but will be in the future. We can however put a condition into the precondition of an action but we have to delegate immediately. So I cannot say: when the Euro rate becomes over 1.4 to the Dollar you become my delegate to sell my house, but I can say: I delegate you now to sell my house on the condition that the Euro rate becomes over 1.4 to the Dollar.Thus according to the Talmud, Figure 58 cannot arise unless $r = t$.

What we learn from the above examples and discussion is the following:

1. We need to specify an annotated path π of delegation.

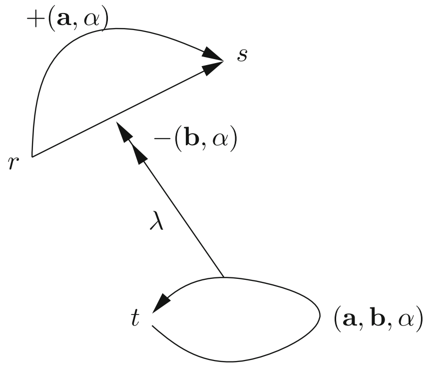


Figure 58.

2. Final revocation or delegation is according to the last delegation (revocation) action in case of switch double arrows but requires a flattening function Λ if we use labels.

3. The double arrows, through their labels, give us who has power over whom to delegate or revoke.

4. We can constrain the movements along the arcs by the geometry of the arcs.

We are devoting special attention to switch labels because this is the simplest model. An agent $\mathbf{a}$ at node t who wants to delegate can go the appropriate arc $t \to t$ with the appropriate double arrow emanating from it. To revoke he just goes again through the arc. To cancel the revocation, he goes again, etc.

This is the simplest model.

Option 4: Reactivity model for the power of attorney *view of Maimonides*

Our starting point for modelling this view is Figure 55, which we modify for representing the case of power of attorney. We get Figure 59.

$\mathbf{a}_1$ can execute action α and move from state t to state s. $\mathbf{b}_1$ cannot do this. In Figure 59, the fact that $\mathbf{a}_1$ can go from t to s is represented by the continuous arrow

$$t \to_{(\mathbf{a}_1, \alpha)} s.$$

The fact that $\mathbf{b}_1$ cannot execute action α and go from t to s is represented by the broken arrow

$$t \not\longrightarrow_{(\mathbf{b}_1, \alpha)} s$$

The big circle around t and the big circle around s are the locations (round tables for t and s respectively) where our agents are sitting ready

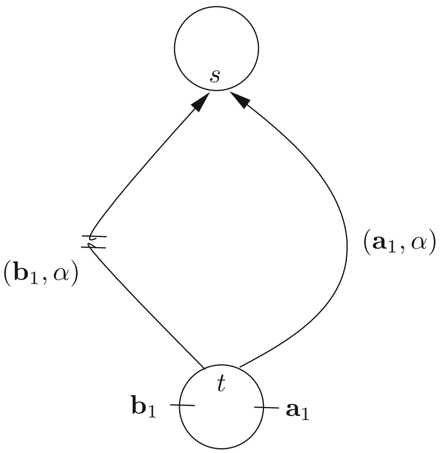


Figure 59.

to delegate and take action. In the *long arm* view, if applied to Figure 59, agent $\mathbf{a}_1$ can delegate (long arm) to agent $\mathbf{b}_1$ the action α by sending a double arrow to the gap of agent $\mathbf{b}_1$ and closing the gap for him so $\mathbf{b}_1$ can move from t to s along his own arrow.

Figure 60 shows this long arm delegation and it should be compared with the slightly different Figure 56.

However, in the case of the *power of attorney* view, $\mathbf{a}_1$ delegates to $\mathbf{b}_1$ by allowing for a double arrow from $\mathbf{b}_1$ to $\mathbf{a}_1$ inside the round talbe circle at t. Figure 61 shows what we mean.

In Figure 61, agent $\mathbf{b}_1$ does have a way to move from t to s executing α. He moves from t to s executing α. He moves along the double arrow $\mathbf{b}_1 \twoheadrightarrow \mathbf{a}_1$ to $\mathbf{a}_1$ position and then moves to s along the $\mathbf{a}_1$ arc

$$t \rightarrow_{(\mathbf{a}_1, \alpha)} s.$$

Here we see how $\mathbf{a}_1$ truly sends $\mathbf{b}_1$ along his own arc!

We note that actually there was no need to draw the broken arc $t \not\longrightarrow_{\mathbf{b}_1, \alpha)} s$ in Figure 59. Since $\mathbf{b}_1$ cannot go from t to s by executing α, it is quite sufficient not to draw an arc for $\mathbf{b}_1$ at all, and the absence of such an arc would indicate that $\mathbf{b}_1$ cannot get to s. However for the sake of comparison, of the *long arm* view with the *power of attorney* view (as shown in Figures 60 and 61) it is advantageous to draw the broken arc in Figure 59.

Note that the Talmudic options for delegation are either the long arm view in all cases, or the *power of attorney* view in all cases. We do not have the mixed option, as in Figure 62. In other words, the Talmud has many cases and debates about delegation. Maimonides takes the view that they

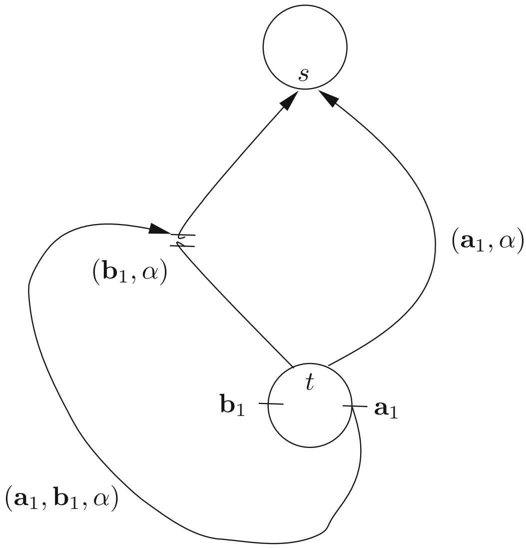


Figure 60. Long arm delegation of α from $\mathbf{a}_1$ to $\mathbf{b}_1$

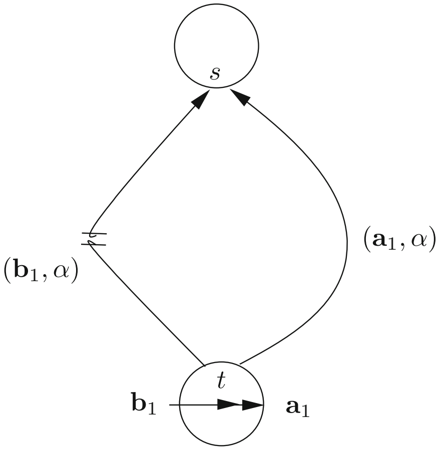


Figure 61. Power of attorney delegation of α from $\mathbf{a}_1$ to $\mathbf{b}_1$

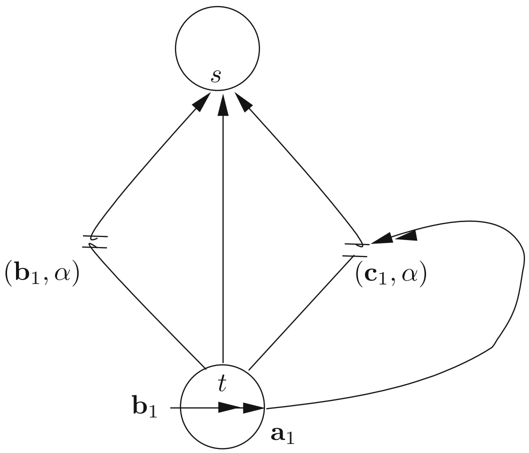


Figure 62.

are all power of attorney cases and the Tur takes the view that they are all long arm cases.

In Figure 62, agent $\mathbf{a}_1$ gives a power of attorney delegation to $\mathbf{b}_1$ to execute α and at the same time gives a long arm delegation to $\mathbf{c}_1$ to execute α.

EXAMPLE 85. To show the difference between the *long arm* and *power of attorney* view, consider the following examples.

1. Ruby and Simon delegate to Levy to dig a hole in the road (to access some pipes). An accident happened. A child fell into the hole and died. Ruby panicked and fled the country.

   According to the *long arm* view of delegation, we consider it as if each of Ruby and Simon dug the hole himself. So each is 100% responsible for damages. According to the *power of attorney* view, Levy was their joint agent and so each is 50% responsible for damages.

2. Another example is Simon delegates Levy to go to Sarah and give her a ring of engagement on behalf of himself. Simon has no money to buy the ring, so Levy (a good friend) buys the ring from his own money and gives it to Sarah.

   The Talmudic argumentation for this story goes as follows:

   (a) According to the *long arm* view, Levy is the long arm of Simon. So it is as if Simon himself gives the ring to Sarah, which he did not buy with his own money. The law says that the ring must be bought with the money of the person to whom Sarah is to be engaged (i.e. Simon). So the engagement action as performed is not valid. According to the *power of attorney* view, the engagement is valid. Levy has a power of attorney to do a job. Levy is not a long arm of Simon. He does the job with his own money. This is fine.

(b) The Talmud could have argued differently from the above.

It could have said that according to the *long arm* view Levy is now Simon and so Levy's money counts as Simon's money, so the engagement is valid. While the *power of attorney* view would say that as an empowered agent, Levy is not Simon, so Levy's money is not Simon's money and so the engagement is not valid.

Our model supports the Talmudic actual argument (a) and not the alternative argument (b) and thus properly models the Talmud.

In Figure 60, it is essentially $\mathbf{a}_1$ which goes through the arc of $\mathbf{b}_1$ to execute α. His long arm $\mathbf{b}_1$ goes for him through the arc

$$t \to_{(\mathbf{b}_1,\alpha)} s$$

which now has no gap.

In Figure 61, $\mathbf{b}_1$ goes thorugh the arc

$$t \to_{(\mathbf{a}_1,\alpha)} s$$

and so $\mathbf{b}_1$ is indeed a power of attorney agent for $\mathbf{a}_1$.

The difference between the cases is which arc is traversed by the agent $\mathbf{b}_1$. Is it $t \to_{(\mathbf{b}_1,\alpha)} s$, (long arm case) or is it $t \to_{(\mathbf{a}_1,\alpha)} s$ (power of attorney case)?

7.5 The reactive switch model for the long arm *view of Tur*

We now present formal definitions for the reactive case with switch double arrows.

DEFINITION 86 (Reactive delegation action model for switch double arrows). Let $\mathbf{A}$ be a finite set of agents and $\mathcal{A}$ a finite set of actions. Let $\mathbb{L}$ be a predicate language and assume that $\{\mathbf{m}\}$ are models for $\mathbb{L}$. We assume the actions $\alpha \in \mathcal{A}$ have preconditions A_α and postconditions B_α in the language $\mathbb{L}$. By a reactive model we mean a tuple $\mathfrak{M} = (S, \mathbf{R}_a, \mathcal{R}, \mathbb{D}, a, \mathbf{m}_t), t \in S$ where

S is a set of states

$\mathbf{R}_a$ is a function giving values in $\{0,1\}$ to tuples of the form $(t, \mathbf{a}, \alpha)$ and to tuples of the form $(t, s, \mathbf{a}, \alpha), t, s \in S, \mathbf{a} \in \mathbf{A}, \alpha \in \mathcal{A}$,

$a \in S$ is the initial state.

$\mathbb{D} \subseteq \mathbf{A} \times \mathcal{A}$ says which agent is a master of which action. Such actions the agent can delegate. We can code $\mathbb{D}$ as part of $\mathbf{R}$ by having

$$\mathbb{D} = \{(\mathbf{a}, \alpha) | \mathbf{R}_a(\mathbf{a}, \alpha) = 1\}.$$

$\mathcal{R}$ is a set of double arrows of the form

$$(t, \mathbf{a}, \mathbf{b}, \alpha) \twoheadrightarrow (u, r, \mathbf{b}, \alpha)$$

or the form

$$(t, \mathbf{a}, \mathbf{b}, \alpha) \twoheadrightarrow (u, \mathbf{b}, \mathbf{c}, \alpha)$$

where $t, u, r \in S, \alpha \in \mathcal{A}, \mathbf{a}, \mathbf{b} \in \mathbf{A}$.

$\mathbf{m}_t$ are models of $\mathbb{L}$.

We assume that if $(t, s, \mathbf{a}, \alpha) \in$ domain $\mathbf{R}_a$ and $(t, s', \mathbf{b}, \alpha) \in$ domain $\mathbf{R}_a$ then $s = s'$.

Note the following:

1. If $\mathbf{R}_a(t, \mathbf{a}, \mathbf{b}, \alpha) = 0$ then at state t, agent $\mathbf{a}$ cannot delegate α as he cannot pass through the arc $t \rightarrow_{(\mathbf{a},\mathbf{b},\alpha)} t$.

2. Whenever $(t, \mathbf{a}, \mathbf{b}, \alpha) \twoheadrightarrow (u, r, \mathbf{b}, \alpha)$ (or respectively, $(t, \mathbf{a}, \mathbf{b}, \alpha) \twoheadrightarrow (u, \mathbf{b}, \mathbf{c}, \alpha)$) is in $\mathcal{R}$ then agent $\mathbf{a}$ by going through the arc $t \rightarrow_{(\mathbf{a},\mathbf{b},\alpha)} t$ (if he is allowed to delegate and other conditions hold) can delegate or revoke the ability of agent $\mathbf{b}$ to execute action α at state u (ending at state v if allowed by other factors), (respectively delegate or revoke the ability of agent $\mathbf{b}$ to delegate or revoke action α at state u to agent $\mathbf{c}$).

DEFINITION 87 (Legitimate path in a model for switch double arrows). Let $\mathfrak{M} = (S, \mathbf{R}_a, \mathcal{R}, \mathbb{D}, a, \mathbf{m}_t), t \in S$ be a model. We define the notion of legitimate annotated path of the form

$$\pi = (a \rightarrow_{e_1} x_1 \rightarrow_{e_2} x_2 \rightarrow \ldots \rightarrow x_{n-1} \rightarrow_{e_n} x_n)$$

where $x_i \in S$ and e_i are annotation labels of the form

$$e_i = (\mathbf{a}_i, \mathbf{b}_i, \alpha).$$

As part of the definition of legitimate path we associate by induction a function $\mathbf{R}_{\pi_i}$ with the initial path $a \rightarrow_{e_1} \rightarrow x_1 \rightarrow \ldots \rightarrow_{e_i} x_i$.

Step 0 $\quad \pi = (a)$ and $\mathbf{R}_{(a)} = \mathbf{R}_a$.

Step $i+1$ Assume we have $\mathbf{R}_{\pi_i}$. We define $\mathbf{R}_{\pi_{i+1}}$, where $\pi_{i+1} = \pi_1 \cup \{x_{i+1} \rightarrow_{e_{i+1}} x_{i+1})$.

Subcase 1 Delegation. In this case $x_i = x_{i+1}$ and $\mathbf{R}_{\pi_i}(x_i, \mathbf{a}_i, \mathbf{b}_i, \alpha) = 1$ and

$$\mathbf{R}_{\pi_{i+1}}(u, v, \mathbf{b}, \alpha) = \begin{cases} 1 - \mathbf{R}_{\pi_i}(u, v, \mathbf{b}_i, \alpha) \\ \text{if } (t, \mathbf{a}, \mathbf{b}_i, \alpha) \twoheadrightarrow (u, v, \mathbf{b}_i, \alpha) \\ \text{is in } \mathcal{R} \\ \text{and } \mathbf{R}_{\pi_i}(u, v, \mathbf{b}_i, \alpha) \text{ otherwise} \end{cases}$$

Similarly

$$\mathbf{R}_{\pi_{i+1}}(u, \mathbf{b}, \alpha) = \begin{cases} 1 - \mathbf{R}_{\pi_i}(u, \mathbf{b}_i, \mathbf{c}_i, \alpha) \text{ if} \\ (t, \mathbf{a}, \mathbf{b}_i, \alpha) \twoheadrightarrow (u, \mathbf{b}_i, \mathbf{c}_i, \alpha) \text{ is in } \mathcal{R} \\ \text{and } \mathbf{R}_{\pi_i}(u, \mathbf{b}_i, \mathbf{c}_i, \alpha) \text{ otherwise.} \end{cases}$$

Subcase 2 Action. In this case $x_i \neq x_{i+1}$ and $\mathbf{R}_{\pi_i}(x_i, x_{i+1}, \mathbf{a}_i, \alpha) = 1$ and $\mathbf{m}_{x_i} \vDash A_\alpha$. Let $\mathbf{R}_{\pi_{i+1}} = \mathbf{R}_{\pi_i}$.

Note that if the action α does not change the state x_i, then we still go to $x_{i+1} \neq x_i$, but we will have $\mathbf{m}_{x_i} = \mathbf{m}_{x_{i+1}}$.

REMARK 88.

1. Now that we have models, we can use modal operators and define satisfaction for them. The only modal operator of interest is the one having to do with delegation.

2. Note that in this model **a** can delegate to **b** to execute action α from state u. This is not general delegation but a very specific one.

DEFINITION 89 (Delegation modalities for switch double arrow). Let us add to the language two modalities $\Diamond$ and $\Diamond_\alpha$, defined as follows in the model $\mathfrak{M}$. $\Diamond$ corresponds to arbitrary legitimate paths. $\Diamond_\alpha$ corresponds to α delegation paths.

1. Let $\pi = a \to_{e_1} x_1 \to \ldots \to_{e_n} x_n$ be an arbitrary legitimate path. Let π' be an extension of π, namely

$$\pi' = a \to_{e_1} x_1 \to \ldots \to_{e_n} x_n \to_{e_{n+1}} y_1 \to \ldots \to_{e_{n+m}} y_m$$

   We say

   - $\pi \vDash \Diamond A$ iff for some extension $\pi', \pi' \vDash A$
   - $\pi \vDash A$ without modalities, iff $\mathbf{m}_{x_n} \vDash A$.

2. A sequence π is said to be α delegation path if for some t we have $x_i = t$ for all i and $e_i = (t, \mathbf{a}_i, \mathbf{b}_i \alpha)$. We can now define

   - $\pi \vDash \Diamond_\alpha A$ iff for some extension π' of π such that $x_n \to_{e_{n+1}} y_1 \to \ldots \to_{e_{n+m}} y_m$ is an α delegation path, we have $\pi' \vDash A$.

3. We can similarly define single modalities for each single connection of the form $t \twoheadrightarrow_{(\mathbf{a},\alpha)} s$. The modality is $\Diamond_{(\mathbf{a},\alpha)}$. Similarly we can define $\Diamond_{(\mathbf{a},\mathbf{b},\alpha)}$.

7.6 Comparison with modern literature

Let us start with the 2001 survey paper [73] and the 2011 logical implementation paper [15] based on it.

Paper [73] addresses an ownership-based framework for access control. Delegation here takes the form of granting access and administrative rights to other agents thus forming chains of granted accesses. The paper is a comprehensive study of the problem of revoking such rights, and of the impact different revocation schemes may have on the delegation chains. Three main revocation characteristics are identified:

a. the extent of the revocation to other grantees (propagation).

b. the effect on other grants to the same grantee (dominance).

c. the permanence of the negation of rights (resilience).

A classification is devised using these three dimensions. The different schemes thus obtained are described, and compared to other models from the literature of the time (up to 2001).

We begin by comparing this scheme with Delegation in the Talmud. The Talmud deals with a person delegating an action to another, to do the action on his behalf. The situation where a delegate can get instructions from two different people to do the same action as a delegate for each of them cannot arise. For example when delegating the selling of a house, only the owner of the house can delegate. If the ownership is shared, one can delegate selling only his share. We cannot have a situation where two different people delegate the same action to a third party. You may ask, how does the Talmud view the delegation of access control? The Talmud does not consider this as a recognised delegation of Halakhic action.[89] According to the Talmud, one cannot delegate an action which does not affect an actual change in some legally recognised state of affairs[90] Selling property, buying, divorcing, getting engaged, are candidates for delegation in the Talmud;

[89] There are references to access permission, but normally as a peripheral sign of a substantive action — such as giving entrance permission or a key as sign of a finalized sale. Access to information is generally restricted in a ruling named after Rabbeinu Gershom (end of the 10th century), appearing in the Responsa of Meir of Rothenburg (1215 - 1293), IV, Ch. 1000 section 22. The general result is that "...the Halachah insists upon the responsibility of each individual not to put himself into a position where he can pry into his neighbour's personal domain, and this responsibility can be enforced by the courts." (Lamm, Norman, "The Fourth Amendment and it's equivalent in the Halachah", Judaism: a quarterly journal, Volume 16, Number 3, Summer 1967, pp. 300–312, p. 303). In modern information-access scenarios (of passcodes and user access), often the question of access to information databases is deliberated along the lines of the authorities given to someone renting a house (Maimonides Rent Laws, 5,5; Responsa Maimonides , 166) — the possibility of subletting, exceeding the capacity, etc.

[90] Sometimes this is called "words cannot be delegated" (Milei lo Mimseran leShaliach). Cf. Tractate Kiddushin 29a, 42a and the commentators.

breaking a window, jumping over a fence — are not. So revocation becomes rather simple here. In the *long arm/extended reach* view of Talmudic delegation, the original master can revoke the last link in the delegation chain and reinstate him at will. The other members of the chain are not part of the *long arm/extended reach* at any stage. In the *power of attorney* view each element in the chain can revoke only the next one. So if the master delegates to an agent and the agent delegates to a subagent, theoretically the master can revoke the delegation to the agent alone, and may be too late if the subagent was already delegated by the agent.

There is one exception: although normally delegation does not work when sending an agent to perform an illegal act,[91] there are rare exceptions[92] In cases of great duress (a coerced agent),[93] the act of murder can be considered as delegated — a person can be considered as an assassin-agent with consequent legal blame on the sender. Theoretically, in the case of murdering somebody, you can have an assassin-agent being delegated to murder by several people. In this case revocation policy is simple. If one of the people revokes the other delegations still stand. Note that according to the *long arm/extended reach* view all masters who delegated to the assassin are each individually murdering the victim (because the assassin is their extension) while with the *power of attorney* view, only the agent-assassin is doing the murdering. The second paper, [15], gives a logical model to the first paper, [73]. The implementation is according to Option 1, where the delegation chains only are modelled as an update system. Recently in paper [18] Barker *et al.* offer a reactive model for access control, which is another way of modelling [73] and more.

Other algorithmic papers dealing with chains of delegations and revoca-

---

[91]The reason often quoted is stated in Tractate Bava Metzia as "Who do you listen to — the master or the student?" (*Divrei Harav V'Divrei Hatalmid divrei Me shomin*). This could mean either: a. that the delegation process did not take off, since the agent could not take the word of the sender over that of the Halakhic prohibition (Joshua ben Alexander HaCohen Falk (1555–1614), Sefer Me'irat Enayim, Choshen Mishpat 182(2), 348(20); Tosafot on Tractate Bava Kama 79a, "Natnu"), or b. that the entire concept of delegation in Halakha is an artificial construct, and thus is simply not defined for actions that are illegal (in the words of A. Kircshenbaum: 'The institute of Shlichut is the creation of the law; as an instrument to break the law — it was not created and does not exist', Dinei Israel Vol. 4, 1973, pp. 55–56).

[92]Three cases are listed in Tractate Kiddushin, 43: the sale of stolen livestock, theft from Hekdesh (consecrated property), and the misuse of property given to the sender for safekeeping.

[93]This is extremely rare. Normally there is harsh moral blame on the person hiring an assassin, but not enforceable in a human court (Maimonides, Laws of Murder and Saving Lives, 2(2-4)). The coerced murder delegation status is considered by some as bona fide delegation (R. Mordechai de Boton in his Responsa, p. 127). For example, the case of David and the death of Uriah in battle (David Kimhi (1160–1235), RaDaK on Shmuel II, 12:9). This understanding is hotly debated, and the issue is thoroughly adumbrated by R. Ovadia Hedaya (1889–1969, prominent Israeli Posek of the previous generation), in his Responsa Yaskil Avdi, Yore De'ah Vol. I, section 6(2).

tion systems are [24; 17; 44; 81; 45; 77; 16; 15].

7.7 Conclusion and discussion

In this paper we gave a preliminary study of delegation in the Talmud and compared it with modern delegation theory. In the Talmud the emphasis is more theoretical; the Talmud is concerned more with the nature of delegation and circumstances for its cancellation, death or madness of the people involved and there is not so much emphasis on revocation. Talmudic delegation is more personal, private persons delegate for the purpose of some legal action, divorce, buying and selling, and so revocation is a simple person to person act. Modern delegation is mainly for access control or the endowment of privileges usually involving large institutions and systems and so revocation protocols and the handling of delegation chains is more central. The emphasis is less conceptual and more algorithmic.

8 QUANTUM STATES AND DISJUNCTIVE ATTACKS IN TALMUDIC LOGIC

This section provides logical modelling for the results contained in the twelfth monograph on Talmudic logic entitled *Fuzzy Logic and Quantum States in Talmudic Reasoning* [11].[94]

This section directly impacts on abstract argumentation theory, temporal and fuzzy arguments and disjunctive collapse. It deals with attacks on a target set of arguments which results in the target to be considered in a quantum like superposition state. The attack is not crisp enough and so cannot be said to be focussed on any individual member or any clear subset of the target. As a result the target set needs to be treated like a quantum superposition of its members.

8.1 Background and orientation

We begin our discussion with several examples.

EXAMPLE 90 (Disjunctive attacks: Story 1). Mr. Smith is a rich old man who wants to donate a very rare classic painting to one of two national museums. He committed the donation in a letter to the two museums

[94] As we have indicated in our first paper and in our book [12] on Talmudic Logic, the aim of this series (of possibly 25-30 books) is twofold:

1. Import logical tools to the service of modelling and explaining Talmudic reasoning and debate.
2. Export ideas and logical constructions from Talmudic debate for the application and use in general logical theory, artificial intelligence and agency and norms.

and copied and approved by Charity Commission, so the donation to one of the two museums was legally done, accepted and in force, except that the choice as to which of the two museums the painting will be given has not been made yet. Mr. Smith said that he would inform the Charity Commission and the museums which museum he would choose in a few days. The donation is in force, however, regardless the status of the choice. Mr. Smith unfortunately died before he made that choice. We are now left with an unclear legal situation regarding ownership. Let a, b and x denoted as follows:

$b =$ the painting does belong to museum b
$a =$ the painting does belong to museum a
$x =$ body of laws regarding ownership.

We have, of course, that a and b are mutually exclusive. Therefore we have that x disjunctively attacks (see [51]) the set $\{a, b\}$. The attack says one of $\{a, b\}$ must be false. We must be clear here.

Suppose we are dealing with n museums . The options are then

$$a_i = \text{the painting belongs to the } i\text{-th museum, } i = 1, ...n.$$

Then we have that x *implies* that exactly on of a_i holds.

Put differently, x implies the set $\{a_i | i = 1, .., n\}$, where the meaning of *imply a set of formulas* is that exactly one of a_i is true, or equivalently that exactly one $\neg a_i$ is false.

Then again reformulating we can say that x *attacks the set* $\{\neg a_i | i = 1, ..., n\}$, where the meaning of *attacking a set* is that exactly one member $\neg a_i$ is false, namely exactly one a_i is true.

In case $n = 2$, we have $a = \neg b$ and $b = \neg a$ and so we have that x *attacks* $\{\neg a, \neg b\}$ is the same as x attacks $\{b, a\} = \{a, b\}$. Talmudic logic debate distinguishes several views on this scenario. The facts on the ground are that the museum's claim that there was a legally binding donation and as for the question of who is beneficiary, a or b, a reasonable deal can be worked out, such as an agreed arrangement of co-ownership, or sharing, or we can let the estate of Mr. Smith continue and choose a museum or we can flip a coin, or ... whatever other symmetrically reasonable solution.

Talmudic logic debate offers two main views on this:

View 1. Quantum like view. This view is that , since Mr. Smith died before making a choice of a museum, ownership is superimposed evenly on both museums, in the same sense as, nowadays, modern quantum mechanics treats the two slits experiment [47]. Recall that in the two slit experiment a single electron is sent towards two slits a and b and the electron passes through both slits as a wave and interferes with itself. So even though logically in classical mechanics the electron is expected to pass only through one slit, it is also a wave according to quantum mechanics and so it passes through both.

The Talmudic debaters holding this view are divided in their verdict:

Option 1. Since a and b are mutually exclusive, there is no longer a donation. The superposition of ownership cancels the donation. The actual Talmudic debate is in connection with marriages but we have adapted the story to Mr Smith and his donation of paintings . See Talmud Bavli, Kidushin, Page 51a and after.

Option 2. The superposition of ownership does not cancel the donation. There is a donation the superposition holds but the fact is that the superposition causes lack of clarity of what to do and it should be undone by court order. The museums should waive their "ownership" back to the estate for otherwise the normal flow of life would be disrupted. After that, the estate can re-donate the painting if they want to.

Of course, those options have implications towards estate tax duties, etc.

Note that both options agree that ownership is super-imposed on both $\{a, b\}$. They differ in their verdict.

View 2. Fuzzy probabilistic view. There is no superposition. There was a donation and we view the scenario as if there was a choice of a museum, except that we do not know what it was, i.e. we treat the case as if Mr. Smith did choose a museum, wrote a letter but died and the letter was lost). So we have a case of purely epistemic uncertainty here and we are expected to provide some mechanism to divide/allocate the painting. For example:

1. Share ownership 50/50.

2. Make a case for one museum over the other, for example, if the painting was in the special area of museum a, then we can argue and reasonably claim that there is high probability that a was chosen.

3. Recommend other arrangements, such as decision by lottery, or time sharing, etc.

Note that there are further implications to View 2. For example if the number of museums involved is very large, we could on probabilistic grounds, agree that the painting remains with the estate, as the probability for each museum to be the owner is very low. This is an interesting Talmudic view. We might think that it is possible for all the museums to form a coalition and ask for the painting. The Talmud will not allow this. To explain this aspect of this view, think of a different scenarios.

Scenario 2.1:(Compare with Example 91). There is one painting which was donated to one museum, from among many paintings and we do not know which one it is. Say both the donator and the museum curator die suddenly. The problem is whether we forbid the estate owners of these paintings to sell any of their paintings for fear that it is the one belonging/donated to the museum. The Talmud view in this case is that since the majority of paintings was not donated we allow the sale.

Compare this scenario with the following variation:

Scenario 2.2. This scenario is the case of donating one painting but not yet deciding which one, and before a decision is made, the owner dies. In this case, the Talmud says that each of the paintings could have been chosen, and so the museum is part owner/potential owner in each painting and so none of them can be sold! This is like modern quantum superposition view.

There are other contexts where this practical probabilistic reasoning makes sense. If one Ebola infected person passed through an airport around the time when there were 2000 others present, we can assume about each of the others that he/she is not infected but cannot treat them as such.

We now conclude our discussion of View 1 and View 2 of the story of Mr. Smith donation of one painting to one of two museums. The main thrust of the story is that there is an attack on the set $\{a, b\}$ without there being any specific attacks on a or on b. The story can continue as follows:

Suppose each of a and of b, independently attacks c, the details of the attack are not important (maybe c is an art critic claiming the painting is a forgery). What is important are the formal options for handling the situation. We have several options for reasoning here

1. c must be out (i.e. false), since either a or b is in (i.e. true) and both attack c.

2. c must be in, since the disjunctive attack is super-imposed on both a and b, so neither is safely to be considered in (true).

3. c is undecided since we do not know exactly what is going on with $\{a, b\}$.

4. c joins $\{a, b\}$ in the status of being a member of the superposition set. In other words, we have that if x disjunctively attacks $\{a, b\}$ and a attacks c, and the attack of x on $\{a, b\}$ is perceived as a superposition on $\{a, b\}$, then the constellation of [x disjunctively attacks $\{a, b\}$ and a attacks c, and the attack of x on $\{a, b\}$ is perceived as a superposition on $\{a, b\}$] is taken to be equivalent to the constellation [x attacks $\{a, b, c\}$ and the attack is perceived as a superposition of $\{a, b, c\}$].

EXAMPLE 91 (Disjunctive attack Story 2). Mr. Smith is a rich man owning 2 original masterpieces. He decides to donate one of these paintings to the museum (a charity). There are several steps to be taken to accomplish this properly. Select the painting, transfer ownership, put conditions on its use and exhibition, get tax relief on the donation, etc., etc.

These steps are persistent in time. Once accomplished they remain so. So the temporal flow is to execute each step properly and then legally end up with the result. The Talmudic scenario is to study, debate and rule in cases where the steps become fuzzy. The question is then to determine what final result we have in this case. The logic behind the Talmudic debate of the various scenarios is the Talmudic fuzzy logic and Talmudic disjunctive attacks.

Scenario 3. Mr. Smith commits a painting to the museum. The museum sends Mr. Jones to go with Mr. Smith to the "storage vault" and choose a painting.

Storyline 1. On the way both die (tragic traffic accident).

Question 1. What does the museum get/own? What would the heirs/estate of Mr. Smith do?

Storyline 2. Mr. Smith and Mr. Jones get to the storage and choose a painting. On the say out of the storage they both die. So we know a painting was chosen but we do not know which one and we have no way of knowing.

Question. Same as before.

Storyline 3. Mr. Smith authorises Mr. Jones to go to the storage and choose a painting. Mr. Jones does that and telephones Mr. Smith and tells him what he chose. A few minutes later Mr. Smith dies of a heart attack and Mr. Jones dies in a tragic accident. The museum knows the government was secretly and unlawfully recording all telephone conversations of prominent citizens. They could try and get the recording of which painting was chosen. This is very difficult because the Government will never admit that it is listening to its citizens. In this scenario, we could find out what painting was chosen, but for all practical purposes, we find ourselves in Storyline 2.

We note that once we are in a state of superposition, like when a painting was donated to one of two museums but not decided which one or one of two paintings was donated to a single museum but not decided which one, we can collapse the superposition retrospectively by, for example, flipping a coin. This is parallel to quantum superposition which can collapse when we do measurements.

Let us now analyse these stories. Let

$$S = \{\pi, \pi'\}$$

be the set of paintings and let $G(x)$ be the predicate that x was given by Mr. Smith to the museum. First we ask: do we know for sure that $\exists x G(x)$ must hold? The problem is that if no painting was chosen, was there a donation?

If we decide that there was a donation, then which painting? Can the museum sell something? Can the museum transfer to another legal entity whatever it has?

Storyline 1. This is the case where a painting was donated but none was chosen. Compare with Example 90.

Rava opinion. There is no deal. The museum gets nothing. (Compare with View 1 of Example 90.)

Abeyei opinion. There was a valid deal. $\exists x G(x)$ is true but we are in doubt as to which painting was given to the museum. We have a case of superposition here.

According to this view, we can recommend some options.

A1: The museum is to give up voluntarily the donation. This is what the law forces them to do.

A2: Alternatively, in practice, they may reach some deal.

1. The estate of Mr. Smith can donate all the paintings to the museum.
2. Choose a paining now.
3. Rotate the donation, rotate every season a different painting.
4. etc.

This may be OK for paintings and museums, but there are other scenarios which are less flexible. Mr. Smith may have two beautiful daughters and he has agreed to give one of his daughters in marriage to Mr. Jones' son. According to Abayei's approach, only option A1 can be taken. No sharing or rotation or anything is possible, only divorce from each of them. According to the law one cannot be married to two sisters at the same time. One cannot even choose one later, because the new choice may not be the correct one and if married to one you cannot have a marital relationship with her sister.

If we look at Storyline 2, here there was a choice of painting or daughter, but we do not know which one. So we can apply a different logical machinery to this case. Maybe we can argue that the museum has all the paintings of Van Gogh except the one which Mr. Smith owns and so it is most likely that the last van Gogh was chosen or in the case of marriage, one can argue that perhaps one of the daughters already knows Mr. Jones' son and the process was most likely aimed at choosing her?

To sharpen the difference between Storyline 1 Abayei and Storyline 2 Abayei, we note the following:

Our storyline 1, Abayei, we accept that $\exists x G(x)$ holds but we do not accept for any $x \in S$ that $G(x)$ holds in a clear cut way, as opposed to some fuzzy way. So Rava says there is no engagement and Abayei says that there is, but it is fuzzy. In Storyline 2, we also accept that for one of $x \in S, G(x)$ holds, but we do not know which one.

We need a logic which can model such distinctions!

8.2 Argumentation networks

We need to model the above examples. We shall use a version of disjunctive argumentation networks [51; 30].

DEFINITION 92. A finite argumentation network has the form (S, R), where S is a finite non-empty set of arguments, and $R \subseteq S \times S$ is an attack relation. We also write $x \twoheadrightarrow y$ in diagrams to express xRy, x attacks y.

EXAMPLE 93. Imagine two pairs of parents planning a joint wedding for their children. They need to compose a list of guests of several types.

1. Relatives from each family
2. Neighbours and friends of parents
3. Friends of the bride and bridegroom
4. Colleagues and co-workers

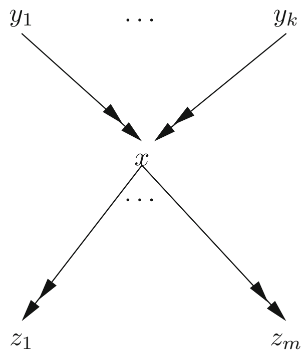
Inviting family can be a problem!

Auntie Bertha might say "I am not coming if that bastard ex-husband of mine is invited". I.e., Bertha $\twoheadrightarrow$ ex-husband.

Grandma Teresa might say "I don't want these kids inviting too many of these hippy crazy friends of theirs, espeically not the drummers". I.e., Teresa $\twoheadrightarrow$ {set of hippies}.

Figure 63 can describe the problematic map which exists:

x is one possible invitee, say Grandma Teresa. She is 109 years old and $y_1, \ldots, y_k$ object to inviting her. Possibly because she is too old and they are worried about her health The reason does not matter. The important



$a \twoheadrightarrow b$ means aRb

Figure 63.

fact here is the double arrow $y_i \twoheadrightarrow x$. This means y_i wants x out. So if y_i is invited, x cannot be invited. Similarly, x objects to $z_1, \dots, z_m$. So the Figure 63 describes the entire configuration around x. We want to define a maximal set E of invited guests such that the following holds:

1. $x, y \in E \Rightarrow x$ does not attack y, (i.e., not xRy). I.e., E is conflict free. No member x of E says "I object" to another member of E.

2. If any x says "why did you invite $z \in$ and you did not invite me? How could you invite this terrible person z"? (i.e., we have $z \twoheadrightarrow z$), then we can say, "we had to invite $y \in E$ and unfortunately, y was against you x" (i.e. for some $y \in E, y \twoheadrightarrow x$).

Such a set E which is also maximal, is called in the argumentation community "a preferred extension". These always exist.

A disjunctive attack has the form $x \twoheadrightarrow H$ where $H \subseteq S$. Its meaning is

- if $x \in E$ then for some $y \in H (y \notin E)$.

This means if you invite x then one of H must not be invited. For example x may be having an affair with both (h_1, h_2). So it is bad taste to invite both. We know about it, but h_1 and h_2 do not know about each other, so it is better not to have them both, says x. We use the notation of Figure 64

DEFINITION 94 (See [51], Definition 3.3).

1. A finite disjunctive argumentation network has the form $\mathcal{A} = (S, \rho)$, where S is a finite set of arguments and $\rho \subseteq S \times (2^S - \varnothing)$, i.e. ρ is a relation of (disjunctive attacks) between elements $x \in S$ and non-empty subsets $H \subseteq S$ denoted as $(x\rho H)$.

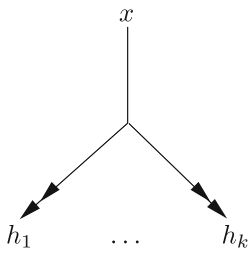


Figure 64.

Let (S, ρ) be a network and let $E \subseteq S$:

(a) We say E is conflict free iff for no $x \in E$ and $H \subseteq E$ do we have $x\rho H$.

(b) We say that E protects α iff for any $z\rho H \cup \{\alpha\}$ there exists a $\beta \in E$ and $E_3 \subseteq E$ and $H_3 \subseteq H$ such that $\beta\rho H \cup H_3 \cup E_3 \cup \{z\}$.

(c) We say E protects itself if it protects each of its members.

(d) We say E is a complete extension if E is conflict free, protects itself and contains all the elements it protects.

Talmudic attack $x\rho H$ wants *exactly one* $y \in H$ to be out. Talmudic logic thinks of it as a collapse of $x\rho H$ to xRy.

The next definition, 95 will explain what we mean by collapse, and give a more correct way to obtain the complete extensions according to Talmudic logic.

8.3 Talmudic argumentation systems

DEFINITION 95. Let $\mathcal{A}$ be a finite disjunctive network and let $x\rho H$ be one of its attacks. We say that a set $\mathbb{F}((x, H))$ is a collapse set for (x, H) if it is the set of all $\mathcal{A}_y, y \in H$ of the form $\mathcal{A}_y = (S, \rho^x_y)$, where $\rho^x_y = (\rho - \{x, H)\}) \cup \{(x, \{y\})\}$. In other words, (S, ρ^x_y) is the network where $x\rho H$ is replaced by $x\rho\{y\}$, i.e. $x\rho H$ collapses to $x\rho\{y\}$.

For each $x\rho H$, let $\mathbf{f}(x, H)$ choose one pair $(x, y), y \in H$. Let $\mathcal{A}_\mathbf{f}$ be the total collapse of $\mathcal{A}$ according to $\mathbf{f}$, defined as $(S, R_\mathbf{f})$, where $R_\mathbf{f} = \{\mathbf{f}(x, H) | x\rho H\}$.

EXAMPLE 96.

1. **Complete collapse**. Consider the network of Figure 65.

   Here we have $x\rho\{a, b\}$ and $y\rho\{b, c\}$. The total collapses are the networks in Figures 66, 67, 68 and 69.

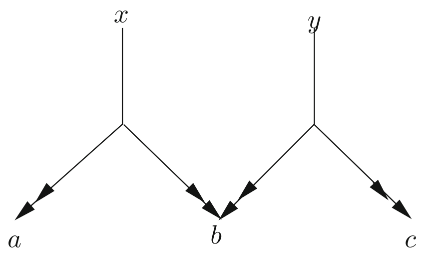


Figure 65.

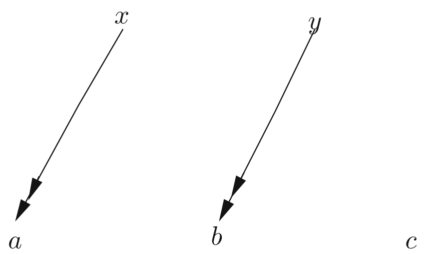


Figure 66.

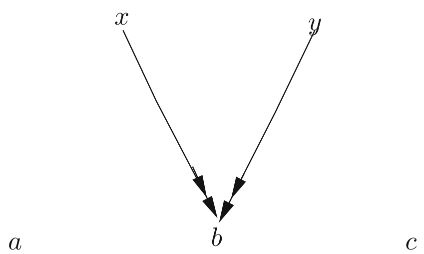


Figure 67.

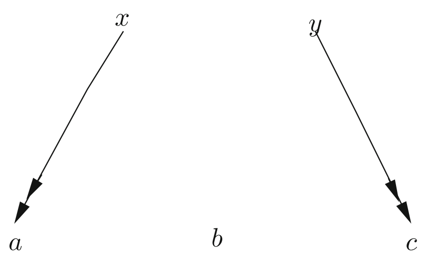


Figure 68.

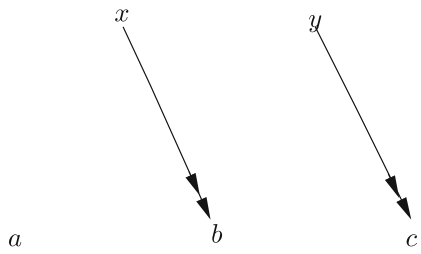


Figure 69.

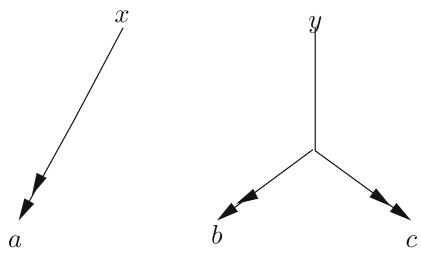


Figure 70.

2. **Partial collapse.** We may have that say $x\rho\{a,b\}$ collapses while $y\rho\{b,c\}$ does not collapse. So we have in this case the possible Figures 70 and 71.

REMARK 97. We have to decide what the Talmud would say about attacks emanating from non-collapesed nodes. Consider Figure 72

In this figure the attack of x on $\{a,b\}$ remains uncollapsed. So this is the final fixed figure. What is our view of $\{a,b\}$? Do we consider them as both in/true (since there is no collapse) for the purpose of the attacks $b \twoheadrightarrow y$, $b \twoheadrightarrow z$ and $a \twoheadrightarrow z$? Or do we regard then as undecided? Do we give them

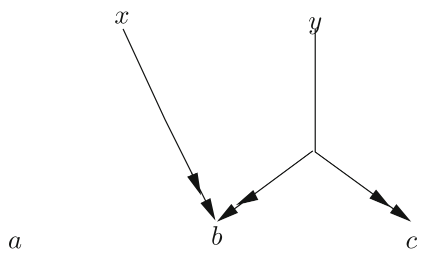


Figure 71.

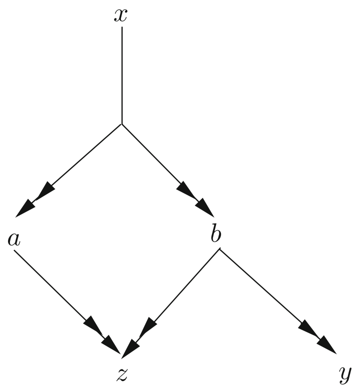


Figure 72.

fuzzy values?

The Talmud approach can be modelled by four values {in, out, undecided, wave}. So we use labelling $x \in \{in, out, und, wave\}$.

So, in Figure 72 we may have that $\{a, b\}$ does not collapse, so we give a, b value "wave" each. This value is passed on to y and z.

If y or z further attack some nodes, they will pass on the value "wave" to their targets.

REMARK 98. Compare with the traditional Caminada labellings and other approaches in [33]. Let us look again at Figure 63 where $y_1, \ldots, y_k$ are all the attackers of x and let us write the conditions on any $\lambda : S \mapsto$ {in, out, und, wave} to be a legitimate Talmudic labelling for a traditional network (S, R) without disjunctive attacks.

(TC1) $\lambda(x) =$ out, if for some $y_i, \lambda(y_i) =$ in.

(TC2) $\lambda(x) =$ in, if for all $y_i, \lambda(y_i) =$ out.

(TC3) $\lambda(x) =$ und, if none of y_i has value $\lambda(y_i) =$ in and some of $\lambda(y_i) =$ und.

(TC4) $\lambda(x) =$ wave, if none of $\lambda(y_i) =$ in and none of $\lambda(y_i)$ is und and some of $\lambda(y_i) =$ wave.

REMARK 99. We now have to define what is a legitimate λ for a network (S, ρ) with disjunctive attacks $\rho \subseteq S \times (2^S - \varnothing)$. We shall reduce this concept by induction to the traditional case with four values as defined in Remark 98. The reduction is by induction on the number of disjunctive attacks in (S, ρ). We first need a concept of constraints on λ.

1. Let (S, R) be an argumentation network of any kind (traditional or Talmudic) with $R \subseteq S \times S$. Let λ_1 be a partial function λ_1 : Subset E of $S \mapsto$ values. We say λ is a legitimate extension under the constraint λ_1 if λ is legitimate and λ agrees with λ_1 on its values.

2. For example in the configuration of Figure 63 we may have the constraint $\lambda_1(y_1) =$ wave. However, if the figure is part of a larger network and y_1 is attacked by a node which needs to be in, then λ cannot overrule λ_1 on the value of y_1.

   When we have a constraint λ_1 it may be the case that no legitimate λ exists with such a constraint.

3. We now define what it means to be a legitimate Talmudic extension for (S, ρ).

   This is done by induction on the number of disjunctive attacks in (S, ρ). We choose a disjunctive attack and do a case analysis of "imaginary" options, (being option (a), (b,i), (b,ii) and (b,iii) below). With each such option we associate a family $\mathbb{F}$ (option) of networks with a lesser number of disjunctive attacks. Each member of each family will yield some legitimate λ by the induction hypothesis, and the totality of these λ are the legitimate extensions for (S, ρ).

   So let us begin:

   Base Case. There are no disjunctive attacks, but there are constraints λ_1, requiring values from $\{\text{in, out, und, wave}\}$. Use principles (TC1)–(TC4) of Remark 98 to get the extensions, if possible.

   Inductive Case. There are disjunctive attacks and there are constraints λ_i. In this case we choose one disjunctive attack. Define the case analysis below and define the sets $\mathbb{F}$ (case number). Any λ found by the inductive hypothesis for any element of these sets will do for our (S, ρ).

   So let us begin the inductive case: Let $x\rho\{h_1, \ldots, h_k\}$ as in Figure 64.

   We distinguish two cases for the Talmudic complete extension λ.

   (a) **case of collapse** In this case the attack of x on $\{h_1, \ldots, h_k\}$ does collapse to one of the attacks $x \twoheadrightarrow h_i$, for some i.

   Therefore we define the legitimate λ for (S, ρ) as any legitimate λ for $\mathbb{F}$ (case (a)) $= \{(S, \rho_i)|$ where $\rho_i = (\rho - \{(x, \{h_1, \ldots, h_k\})) \cup \{(x, \{h_i\})\}\}$ respecting the constraints λ_1.

   (b) **case of no collapse** In this case we distinguish three cases.

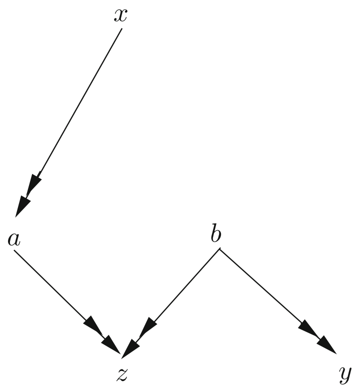


Figure 73.

i. x is out. In this case there is no attack and we let the legitimate λ for (S, ρ) to be any one of the legitimate λ of $\mathbb{F}$ (case (b,i)) $= \{(S, \rho_i)$ of case (a) but with the additional constraint to λ_1 being the constraint $x =$ out$\}$.

ii. x is in or $x =$ wave. In this case there is no collapse and so we have the additional constraints for λ_1 being $h_1 = h_2 = \ldots = h_k =$ wave. So we let the legitimate λ for this case for (S, ρ) to be any legitimate λ for the network $\mathbb{F}$ (case (b,ii)) $= \{(S, \rho')$ where $\rho' = \rho - \{(x, \{h_i, \ldots, h_k\})\}$ under the constraint λ_1 augmented by the additional constraints $x =$ in or $x =$ wave ,respectively and $h_i =$ wave for $i = 1, \ldots, k\}$.

iii. x is und. In this case we look at (S, ρ') as in case (ii), with the additional constraints to λ_1 being the constraint $x = h_1 = \ldots = h_k =$ und.

EXAMPLE 100. Let us see what the Talmud would do with Figure 72.

Here we have only one disjunctive attack $x\rho\{a, b\}$ for which we know $x =$ in because x is not attacked. So there are two possibilities for this attack.

1. The attack collapses and so $x \twoheadrightarrow \{a, b\}$ is to be replaced either by $x \twoheadrightarrow a$ or by $x \twoheadrightarrow b$, giving rise to Figure 73 or Figure 74.

2. The attack does not collapse, giving rise to Figure 75 with the constraints shown.

So the possible extensions according to Remark 98 are:

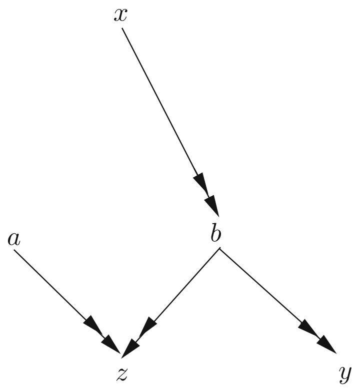


Figure 74.

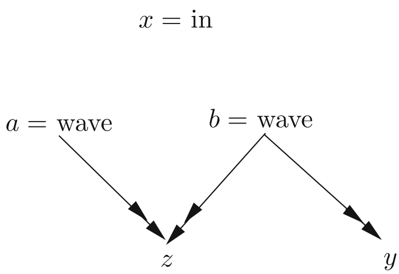


Figure 75.

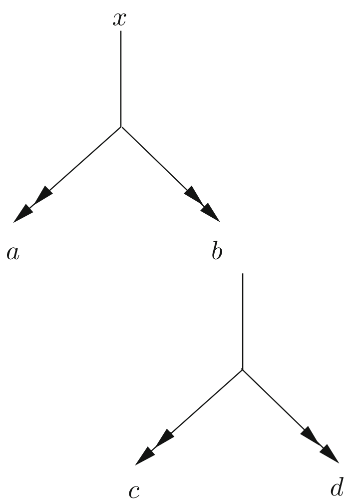


Figure 76.

$$\begin{aligned}
&\lambda_1 : x = \text{ in}, \ a = \text{ out}, b = \text{ in}, z = y = \text{ out} \\
&\lambda_2 : x = \text{ in}, \ a = \text{ in}, b = \text{ out}, z = \text{ out}, y = \text{ in} \\
&\lambda_3 : x = \text{ in}, \ a = b = z = y = \text{ wave.}
\end{aligned}$$

EXAMPLE 101. Consider the network of Figure 76.

Let us agree that $x\rho\{a,b\}$ does collapse while $b\rho\{c,d\}$ does not collapse.

The extensions are the following, calculated intuitively.

$$\begin{aligned}
&\lambda_1 : x = \text{in}, \ a = \text{ out}, \ b = \text{ in}, \ c = d = \text{ wave} \\
&\lambda_2 : x = \text{in}, \ a = \text{ in}, \ b = \text{ out}, \ c = d = \text{ in}
\end{aligned}$$

Let us now follow our inductive procedure of Remark 99 and let us start inductively from $b\rho\{c,d\}$. We get four options, as seen in Figures 77, 78, 79 and 80. The constraints are written in the figures.

For each of the Figures 77–80 we deal with the attack $x\rho\{a,b\}$. These split into two figures each. One with the attack of x on a and one with the attack of x on b.

Some of these will not be possible.

Here are the Figures:

We see that the inductive procedure gave us λ_1 and λ_2 as we expected.

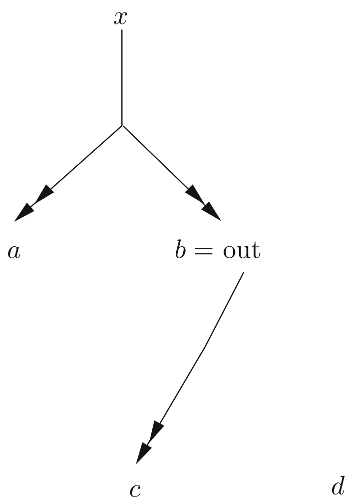


Figure 77.

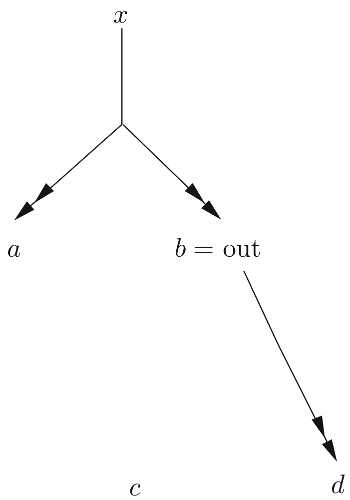


Figure 78.

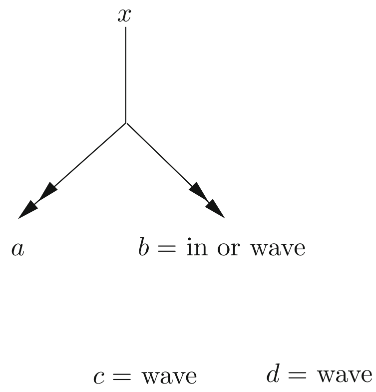


Figure 79.

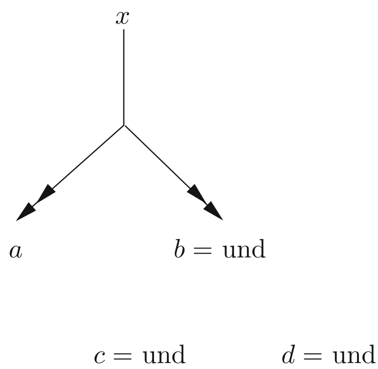


Figure 80.

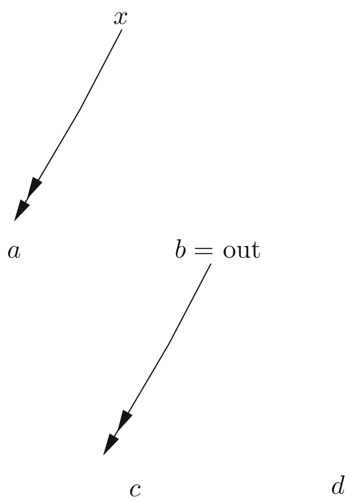


Figure 81. Not possible.

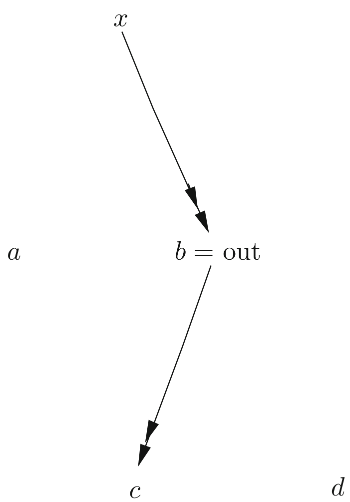


Figure 82. Possible, gives λ_2.

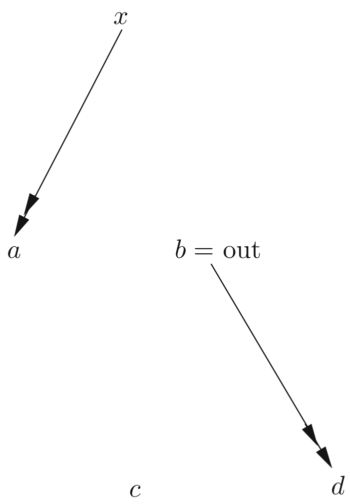


Figure 83. Not possible.

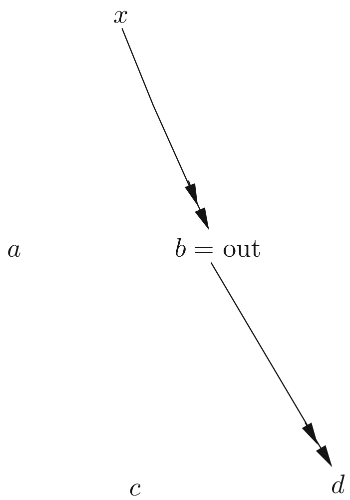


Figure 84. Possible, gives λ_2.

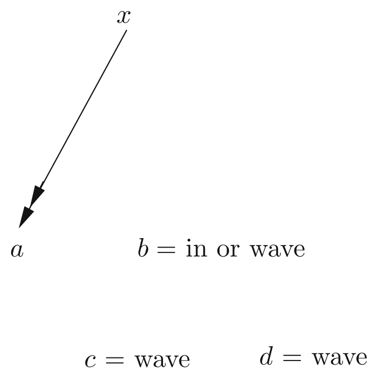


Figure 85. Possible only with $x =$ in. Gives λ_1.

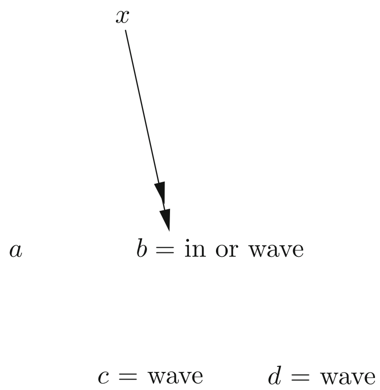


Figure 86. Not possible

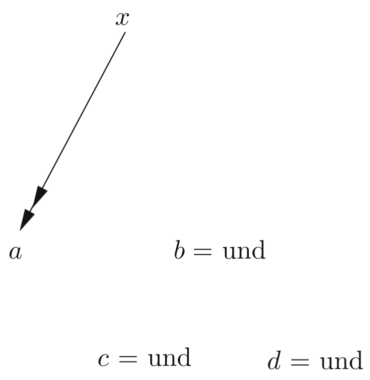


Figure 87. Not possible.

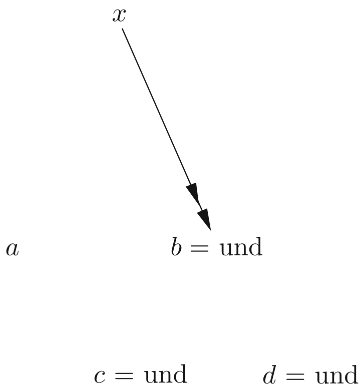


Figure 88. Not possible

8.4 Using Bochman's collective argumentation

In his paper [30], Bochman considered conjunctive disjunctive attacks of the form $G \twoheadrightarrow H$, where both G and H are subsets of S. The intended meaning of $G \twoheadrightarrow H$ is that if all embers of G are in, then at least one member of H is out. The treatment of this notion is straightforward (see [51] and Bochman [30]) using axiomatic properties on $G \twoheadrightarrow H$ to characterise various types of semantics. For example, he considered the following axioms:

Montonicity. If $G \twoheadrightarrow H$ then $G \cup G' \twoheadrightarrow H \cup H'$.

Symmetry. If $G \twoheadrightarrow H_1 \cup H_2$ then $G \cup H_1 \twoheadrightarrow H_2$.

Affirmativity. The empty set is not attacked.

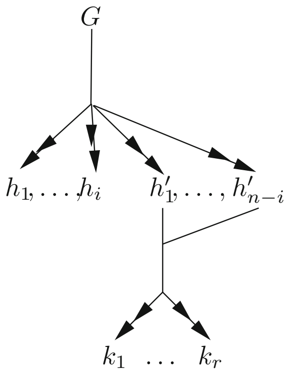
Locality. If $G \twoheadrightarrow H_1 \cup H_2$ then $G \twoheadrightarrow H_1$ or $G \twoheadrightarrow H_2$.

Using this axiomatic approach, we want to characterise Talmudic disjunctive attacks. Let us list the properties we need to characterise:

(P1) If $G \twoheadrightarrow H$ then either (a) or (b) holds

- (a) For exactly one $y \in H$ we have $G \twoheadrightarrow \{y\}$. (This is collapse.)
- (b) For none of $H' \subsetneq H$ do we have $G \twoheadrightarrow H'$. This is quantum superposition.)

(P2) If $G \twoheadrightarrow H$ is a quantum superposition attack on H and for some $H' \neq \varnothing$ such that $H' \subseteq H$ we have that $H' \twoheadrightarrow K$ then $H' \twoheadrightarrow K$ is also a quantum superposition attack and furthermore $G \twoheadrightarrow (H - H') \cup K$ also holds as a quantum superposition attack.



If the value of h'_j is "wave" because of G then the values of $k_1, \ldots, k_r$ is also "wave".

Figure 89.

Figure 89 explains the idea in terms of labels. Any attacker with a wave label propagates this label to its targets.

Note that (P2) can be better understood in positive terms. If $G \to H$ and $H' \to K$ and $H' \subseteq H$, then $G \to (H - H') \cup K$.

The following Bochman-style axioms can characterise (P1) and (P2).

(AP1) If $G \twoheadrightarrow H$ then $H \neq \varnothing$ and

$$\bigvee_{y \in H} \{[G \twoheadrightarrow \{y\}] \wedge \bigwedge_{z \neq y} \neg G \twoheadrightarrow \{z\}\} \vee (\bigwedge_{H' \subsetneq H} (\neg(G \twoheadrightarrow H')))$$

(AP2) If $G \twoheadrightarrow H$ and $\varnothing \neq H' \subsetneq H$ and $H' \twoheadrightarrow K$ then $G \twoheadrightarrow (H - H') \cup K$ and $\bigwedge_{K' \subsetneq K}(\neg(H' \twoheadrightarrow K'))$ and $\bigwedge_{K' \subsetneq (H-H') \cup K}(\neg(G \twoheadrightarrow K'))$.

8.5 Conclusion and discussion

We saw that Talmudic disjunctive attacks require four values, {in, out, und, wave} and differs from [51] in two senses:

1. The attack on a target set H can turn the target set into having the value "wave" for quantum like superposition. This value is then propagated further by the members of the target set when they attack further targets.

2. When attacking a target set H the attack can collapse to attacking a single $y \in H$. Note and compare that the disjunctive attacks in general can collapse to attacking a subset $\varnothing \neq G \subseteq H$.

In our case the options for the attack on a set H is either an attack on a single member of H or rendering H into a wave quantum superposition state.

3. We note here, that in view of references [51] and [30], one of our referees commented as follows:

   > "Two things should be distinguished here. First, the quantum superposition idea, taken by itself, is more or less comprehensible and internally coherent. The only question is whether it is relevant to argumentation, and it is here that I have my reservations. An argument may collectively attack a set of arguments H just because it disproves one of the joint conclusions of H. This is not a quantum phenomenon, and different arguments in H can still be separated by other arguments, and by different attacks they (separately) create against further arguments. So, if your interpretation insists on existence of a special 'superposition' of a set of arguments, it ought to be represented as an entirely new connective for combining arguments, over and above the existing argumentation machinery."

   We comment in return that adding a connective on a set H which turns it into a quantum state is too strong a move. In our system a set H may turn into a quantum state only when attacked. It is the nature of the attack that causes the quantum state. We can certainly investigate a connective, say $\mathbb{Q}(H)$, which turn H into a quantum state but it will be different

   (a) A disjunctive attack cannot turn a single point set into a quantum state, but the connective $\mathbb{Q}$ can do that (and make this point propagate the "wave" value).

   (b) Adding $\mathbb{Q}$ yields a different logic. (Certainly worthy of investigation.)

4. Another referee brought to our attention the works of Andrew Schumann, [104; 105], connecting Talmudic Logic to parallel computation. The referee spent a lot of effort going through our papers commenting how the quantum view can, and maybe should, be replaced by a parallel computation view.

   Let us respond to the referee's proposal, i.e. comment on the connection between the phenomena that we describe and parallel computation. Parallel computation describes computational processes that are carried out in parallel. It focuses on processes that cannot be done serially. The focus of the logical problems in our case (a man that marries one of two women) is not connected to the serial question. If the problem was that one cannot decide the state of the one without previously knowing the state of the other or vice versa, the question of serialism would have been relevant. But our problem is totally different. The state cannot be fully decided, even if we do the computation serially. The fact that one of the women is married prevents the marriage of the other, without any connection to the order of computation. It is therefore a problem of Quantum Logic (intertwining of states) and not Parallel Computation Logic. Said another way, in our case there is a complex interaction between the two channels of the problem (like the interaction between distant particles in an ERP experiment). This is the focus of our investigations, and not just the existence of two parallel channels. The logic of the created state is what we discuss, and this logic is Quantum Logic. We have no interest in how to do the computation that helps us reach the conclusion that this is indeed the state.

9 IDENTITY MERGING AND IDENTITY REVISION IN TALMUDIC LOGIC

Background and orientation

This section makes two statements about AGM revision Theory:

1. AGM deals with abstract revision, where the theories $\mathbf{T}$ we deal with are purely logical and symbolic and without any accompanying interpretation in an application area. So the revision postulates and machinery we use are mathematical and have no guidance from an application area. In practice, theories $\mathbf{T}$ are formulated with an area in mind and when they need to be revised, we can fall back on the application area to guide us on how to revise them

2. We illustrate our point by looking at Identity Merging theories in Talmudic Logic and examining how Talmudic logic does the revision in this case. Talmudic logic is a typical case of a hierarchy of theories where higher nodes in the hierarchy guide us on how to revise the lower nodes.

Let us start. Suppose we are given a monadic theory $\mathbf{T}$ about the free variables x and y. If x and y are never quantified upon in $\mathbf{T}$ they can also be viewed as constants x and y. So $\mathbf{T}$ is built up in classical logic from monadic predicates $\{P_1, P_2, \ldots\}$ and the classical connectives $\{\neg, \wedge, \vee, \rightarrow\}$ and the quantifiers $\forall, \exists$ and possibly the equality symbol $=$.

For example the theory $\mathbf{T}$ may have in it $P(x)$ and $\neg P(y)$.

We now add to the theory the revision input $x = y$. The new theory is $\mathbf{T}_{x=y}$, $\mathbf{T} \cup \{x = y\}$ is inconsistent. We need a belief revision mechanism to revise $\mathbf{T}$ so that it is consistent with the input. We can of course use one of the many AGM approaches and algorithms for this case, but these are too general and we need a more specialised tailored approach for our case. This is a very specific form of input and belief revision, of the form which we are calling "identity merging". We have knowledge about two distinct individuals (so we believe) and we discover that they are the same individuals. Now we have to reconcile what we know. Another very common case is where two conflicting bodies of laws apply to the same individual case and we need to decide how to proceed. Such cases require specialised procedures possibly tailored for each application area (case study). So formally we have a theory $\mathbf{T}$ and two distinct constants or variables x and y and we add the input $x = y$. We need not necessarily deal with a language with identity. If we do not have "=" in the language, we can still take $\mathbf{T}$ and take a new variable z and substitute in $\mathbf{T}$ the variable z for every free occurrence of x or of y. We will get a new theory which we can denote by $\mathbf{T}(x = y = z)$ which is inconsistent (containing $P(z)$ and $\neg P(z)$) and in need for revision. Note on the notation side, we can regard x and y either as constants or as free variables, this does not matter as long as we do not apply any quantifiers to them. So in the sequel we might talk about constants a or b or variables x and y. The choice depends on stylistic reasons.

We have three options here for revision.

1. Use the well known machinery, see, for example [62], which is this case means we choose one of each atomic contradicting pair $\{P(z), \neg P(z)\}$;

2. Use some new approach taking advantage of the specific form of the revision problem.

3. Use the Talmudic Logic approach;

Let us give some examples before we continue.

EXAMPLE 102. Consider a university system with a Rector x and Head of Department of Informatics y. The university has regulations which say among others that:

1. The Rector can offer a position to a candidate and this is legally binding.

2. A Head of Department can offer a position to a candidate (in his department) but it is not legally binding, but is subject to approval by the Rector. The Head of Department should use a standard letter form which makes this clear.

Suppose now that there is a big fight between the Head of Department and his professors and the Head resigns and there is no agreement about a successor. Someone has to run the day-to-day matters of the Department, and so the Rector becomes acting Head of this Department for the time being. The Rector in his capacity as Head, offers a position in the Department to a candidate c. The standard letter which one sends in such a case says that this offer still requires the approval of the Rector but that the Department and its Head are confident that the Rector will approve the offer.

In this case the Head, who writes the letter as Head, is also the Rector, who needs to approve the appointment. The question is:

> Is this letter binding or not?[95]

We have:

> Rector writes $\to$ binding
> Heads writes $\to \neg$ binding

If we revise by the input Rector = Head, do we take binding or $\neg$ binding?

Commonsense says that this is a binding offer.

EXAMPLE 103. This is a real example recently discussed in the American press. It relates to the Boston terrorist bombing.[96] The terrorists were American citizens and so there were two options:

1. Viewed as terrorists, send them to military trial or even to Guantanamo Bay detention camp.

2. Treat them as American citizens and send them through the American legal system.

In principle what is happening here is that we have two bodies of laws and regulations:

> $\mathbf{T}_1(a)$ = Rules for a, a typical terrorist
> $\mathbf{T}_2(b)$ = Rules for b, a typical US citizen.

The theory is $\mathbf{T} = \mathbf{T}_1(a) \cup \mathbf{T}_2(b)$. The input, forced upon us by the real world, is $a = b$.

[95] This actually happened to D. Gabbay in 1972. The perceptive reader might wonder why the Rector was ambiguous in his letter? Well, it may have been deliberate or he may have wanted to follow clear procedural lines and first write as Head and follow it up as Rector. There is more to it than that. Can the recipient of the letter assume that for all practical purposes he/she actually got the job or is there still a practical possibility that the Rector would say "I approve the appointment as Head of Department but I do not approve it as Rector"?

[96] https://en.wikipedia.org/wiki/Boston_Marathon_bombing, http://www.britannica.com/event/Boston-Marathon-bombing-of-2013.

The aim of this section is to formalise and introduce the Talmudic approach. The approach is general and can be used in AI for this case, as an alternative methodology to AGM.

The AGM approach would simply take out from $\mathbf{T}$ one of $\{P(z), \neg P(z)\}$ and restore consistency (assuming P is atomic).

The Talmudic approach will do something different. To introduce it, however, we begin with describing an intermediate non-monotonic approach which is not the Talmudic one, but has an independent interest and would lead into the Talmudic approach.

The non-monotonic approach (ANM vs. AGM) says that the language of $\mathbf{T}$ (i.e. $P_1, P_2, P_3, \ldots$) is only a surface language $\mathbb{M}$ and the fact that $\mathbf{T}$ contains $P(x)$ and $\neg P(y)$ stems from deeper level non-monotonic considerations in a deeper language $\mathbb{L}$. When we revise with $x = y = z$, we have to go to the deep level non-monotonic theory governing P, x and y and see what happens there and then decide whether to contract $P(x)$ or to contract $\neg P(y)$. Thus the non-monotonic approach is a refinement of the AGM approach, where the choice of which of the contradicting pair $\{P(z), \neg P(z)\}$ to take out is made using the extra non-monotonic machinery available in the system. This is best understood when actually defined. Let us propose an ANM model.

DEFINITION 104.

1. Let $\mathbb{M}$ be the monadic classical predicate language with unary predicates $\{P_1, P_2, \ldots\}$ and variables and constants $\{x, y, z, c_1, c_2, \ldots\}$. We say that $\{P_i, c_j\}$ are the predicates and constants of $\mathbb{M}$. Let $\mathbb{L}$ be an expansion of $\mathbb{M}$ with additional predicates and constants $\{A, B, \ldots\}$ and and $\{d_1, d_2, ...\}$.

2. With each constant c and predicate P of $\mathbb{M}$ we associate a family of predicates and constants from $\mathbb{L}$ (which include P and c). Let us use the notation $\mathbb{F}(P, c)$ for this family. For example, let

$$\begin{aligned} \mathbb{F}(P, a) &= \{A\} \cup \{P, a\} \\ \mathbb{F}(P, b) &= \{B\} \cup \{P, b\}. \end{aligned}$$

   (We will not repeat "$\{P, c\}$" any more.)

3. Assume that we have a non-monotonic consequence $\Vdash$ governing the language $\mathbb{L}$ and a theory Δ of facts for the new predicates $\{A, B \ldots\}$ of $\mathbb{L}$ which contains $\mathbb{M}$.

4. Assume that our surface theory $\mathbf{T}$ is the result of Δ. Namely

$$P(x) \in \mathbf{T}_{\Delta, \mathbb{F}} \text{ iff } \Delta \upharpoonright \mathbb{F}(P, x) \Vdash P(x).$$

5. We say that $\mathbf{T} = \mathbf{T}_{\Delta, \mathbb{F}}$ is derived from Δ using $\Vdash$.

EXAMPLE 105. Consider the surface predicates and constants of $\mathbf{T}$ to be P, a, b.

Let

$$\mathbb{F}(P, a) = \{A\}$$
$$\mathbb{F}(P, b) = \{B\}$$

Assume our non-monotonic logic for $\mathbb{L}$ relies on more specificity and that we have a theory Δ with the following rules:

1. $A(a) \to P(a)$
2. $B(b) \to \neg P(b)$
3. $A(b) \wedge B(b) \to P(b)$
4. $A(a)$
5. $B(b)$

Our theory $\mathbf{T}$ will contain therefore $P(a)$ and $\neg P(b)$. This is because showing $P(a)$ we can use only clauses 1. and 4. and showing $P(b)$ we can use only clauses 2. and 5.

Now let us see what happens if we add the input $a = b$. This changes the language we consider from the separate $\mathbb{F}(P, a)$ and $\mathbb{F}(P, b)$ into the joint $\mathbb{F}' = \mathbb{F}(P, a) \cup \mathbb{F}(P, b)$. Now the clauses to consider from Δ are 1. to 5.

But now, because of more specificity

$$\Delta \upharpoonright \mathbb{F}' \Vdash P(b)$$

and so we revise $\mathbf{T}$ in view of the input $a = b$ by contracting $\neg P(b)$. Thus we see that whereas ordinary AGM theory allowed for arbitrary choice in either contracting $P(b)$ or contracting $\neg P(b)$, the non-monotonic background theory Δ, recommended contracting $\neg P(b)$. Intuitively we asked ourselves (and asked Δ) where do $\mathbf{T} \vdash P(a)$ and $\mathbf{T} \vdash \neg P(b)$ come from and we made a decision based on the answer.

We realise that perhaps the non-monotonic system may not resolve the issue. We can rely on another level (i.e. another Δ' related to Δ in a similar way to the relation of Δ to $\mathbf{T}$) and language to resolve the issue. The details are not so important as the overall approach.

REMARK 106. The perceptive reader might ask: "Where does the richer language theory Δ come from? Isn't it a bit artificial, to introduce it just to solve the problem, like a special distance/choice function for AGM revision?

I see how it works and where it comes from in the examples, but in the general case?".

Our answer to this is that indeed laws and regulations come from practical situations where in the background there are undesirable cases to be avoided.

So for each specific $\mathbf{T}$ tailored for a specific application area there will be a corresponding Δ. For a general theory we must stipulate and study a general recursive hierarchy of $\mathbf{T}_n$, where $\mathbf{T}_{n+1}$ acts as the "Δ" of $\mathbf{T}_n$.

Let us now work towards giving a complete formal presentation of the above ideas.

DEFINITION 107. Let $\mathbb{L}$ be a language containing $\neg$ and let $(\Delta, \Vdash)$ be a non-monotonic theory in $\mathbb{L}$. This means that Δ is a set of formulas of $\mathbb{L}$ and $\Vdash$ is a consequence relation of the form

$$A_1, \ldots, A_n \Vdash B$$

where A_i, B are formulas of $\mathbb{L}$ and $\Vdash$ satisfies the following conditions:

1. Reflexivity
$$A_1, \ldots, A_n \Vdash B, \text{ if } B \in \{A_i\}$$

2. Cut
$$A_1, \ldots, A_n X \Vdash B$$
and
$$A_1, \ldots, A_n \Vdash X$$
imply
$$A_1, \ldots, A_n \Vdash B$$

3. Restricted monotonicity
$$A_1, \ldots, A_n \Vdash B$$
and
$$A_1, \ldots, A_n \Vdash C$$
imply
$$A_1, \ldots, A_n, B \Vdash C.$$

4. Let $\theta', \theta \subseteq \Delta$ be two finite subsets of Δ.

   We may have $\theta \Vdash B$ but $\theta \cup \theta' \Vdash \neg B$ or alternatively $\theta \Vdash \neg B$ but $\theta \cup \theta' \Vdash B$.

5. $(\Delta, \Vdash)$ is said to be consistent iff for no $B, \theta \subseteq \Delta, \theta$ finite, we have $\theta \Vdash B$ and $\theta \Vdash \neg B$.

6. θ is said to *decide* a wff B if we have either $\theta \Vdash B$ or $\theta \Vdash \neg B$ (as opposed to neither).

7. θ is said to be minimal theory deciding B, if θ decides B and no $\theta' \subsetneqq \theta$ decides B.

8. We say that $\Delta \Vdash B$ iff there is a $\theta \subseteq \Delta$, θ finite, which decides B and for every minimal θ which decides B we have $\theta \Vdash B$.

DEFINITION 108.

1. Let $\mathbf{T}$ be a classical consistent set of wff, (considered as a theory) in a language $\mathbb{M}$. Let Δ be a consistent non-monotonic set of wffs (considered a theory) in a richer language $\mathbb{L} \supseteq \mathbb{M}$. Let $\Vdash$ be its consequence relation. Let P be a unary atomic predicate of $\mathbb{M}$ and let $a_i, i = 1, \ldots, k$ be distinct constants of $\mathbb{M}$. Let $\mathbb{F}$ be a function giving for each $\alpha = \{P, a_1, \ldots, a_k\}$ a sublanguage $\mathbb{F}(\alpha)$ of $\mathbb{L}$. We assume that $\alpha \subseteq \mathbb{F}(\alpha)$. Note that we may have
$$\Delta \upharpoonright \mathbb{F}(\alpha) \Vdash \pm P(a)$$
but
$$\Delta \upharpoonright \mathbb{F}(\beta) \Vdash \mp P(a),$$
for $\beta \supsetneqq \alpha$. This can happen because $\Vdash$ is non-monotonic.

2. Let $\mathbf{T}$ and Δ be as in (1) above. We say that Δ supports $\mathbf{T}$ if the following (*) holds for each unary P and constant a:

(*) $\mathbf{T} \vdash P(a)$ iff $\Delta \upharpoonright \mathbb{F}(P, a) \Vdash P(a)$.

3. Let a, b be two distinct constants of $\mathbf{T}$. Denote by $\mathbf{T}_{a/b}$ the theory obtained from $\mathbf{T}$ by replacing every occurrence of a by the constant b.

4. Similarly, let P, Q be two unary predicates, we let $\mathbf{T}_{P/Q}$ be the theory obtained by replacing every occurrence of P by Q.

REMARK 109. Let $\mathbf{T}$ be a monadic theory with monadic predicates $\{P_1, \ldots, P_k\}$ and constants $\{c_1, \ldots, c_m\}$. Let $\mathbf{m}$ be a classical model of the language with $\{P_j, c_i, \neg, \wedge, \vee, \rightarrow, \forall, \exists, =\}$. For this language, any model $\mathbf{m}$ has to say the following

1. For each c_i and P_j $\mathbf{m}$ has to say whether $P_j(c_i)$ or $\neg P_j(c_i)$ holds.

2. For each vector ε of $\{0, 1\}$ values of length k, let
$$\alpha_\varepsilon(x) = \bigwedge_{j=1}^{k} P_j^{\varepsilon_j}(x)$$
where $P_j^1(x) = P_j(x)$ and $P_j^0(x) = \neg P_j(x)$, and x is a variable. Then the model $\mathbf{m}$ has to say whether $\exists x \alpha_\varepsilon(x)$ holds for each ε and, if equality is in the language, $\mathbf{m}$ has to say how many different such elements exist, $0, 1, 2 \ldots$ or infinity, i.e. for each n, $\mathbf{m}$ must say whether $\exists x_1, \ldots, x_n \bigwedge_{i \neq j} x_i \neq x_j$.

If there is no equality in the language then a model $\mathbf{m}$ for the language can be characterised by a wff $\varphi_{\mathbf{m}}$ of the form

$$\varphi_{\mathbf{m}} = \bigwedge_{\varepsilon} \pm\exists x \alpha_{\varepsilon}(x) \wedge \bigwedge_{i,\varepsilon} \pm\alpha_{\varepsilon}(c_i).$$

Let us assume we do not have equality.

Then we can assume that any $\mathbf{T}$ has only a finite number of finite models.

REMARK 110. Let $\mathbf{T}$ be a monadic theory without equality as in Remark 109. Let $\mathbf{m}$ be a model for the language of the form $\varphi_{\mathbf{m}}$ as in Remark 109.

We now investigate the effect of the additional revision/input information that c_1 and c_2 are the same (i.e. $c_1 = c_2$). We ask whether $\varphi_{\mathbf{m}}$ is still consistent under the substitution of c_2 for c_1 (or $c_1 = c_2$)?

This depends on what conjuncts appear in $\varphi_{\mathbf{m}}$. The critical ones to be watched are triples $\varepsilon, \varepsilon', P$ such that

$$\varphi_{\mathbf{m}} \vdash \alpha_{\varepsilon}(c_1) \wedge \alpha_{\varepsilon'}(c_2)$$

and such that

$$\alpha_{\varepsilon}(c_1) \vdash P(c_1) \text{ and } \alpha_{\varepsilon'}(c_2) \vdash \neg P(c_2).$$

In other words, the problem is that the model $\mathbf{m}$ says for some set of unary predicates $\{P_{i_1}, \ldots, P_{i_n}\}$ the opposing pair $\{\pm P_{i_r}(c_1) \text{ and } \mp P_{i_r}(c_2)\}$.

Since we are claiming $c_1 = c_2$, we need to choose only one of them, if we want to maintain consistency.

We have a similar problem if we input equality of two predicates, say $P_1 = P_2$. There may be some $c_{j_1}, \ldots, c_{j_n}$, which the model says

$$\pm P_1(c_{j_r}) \text{ and } \mp P_2(c_{j_r})$$

again, we have opposing pairs and again we need to choose one of them if we want to maintain consistency.

Our theory of identity merging will tell us how to choose one from each opposing pair and thus maintain consistency. Our identity merging theory is a refinement of AGM for this particular case. AGM does not care how we choose.

DEFINITION 111. Let $\mathbf{T}$ be a complete and consistent monadic theory with constants $\{c_i\}$ and unary predicates $\{P_j\}$, and let $(\Delta, \Vdash)$ be a supporting theory for $\mathbf{T}$ as in Definition 108. Let a, b be two distinct constants and consider $\mathbf{T}_{a/b}$ and assume that it is inconsistent. The merge revision of $\mathbf{T}_{a/b}$ is performed as follows.

Since $\mathbf{T}$ is complete, $\mathbf{T}_{a/b}$ being inconsistent means that either $\mathbf{T} \vdash P(a) \wedge \neg P(b)$ or that $\mathbf{T} \vdash \neg P(a) \wedge P(b)$ for some predicates P.

Assume without loss of generality that the former holds. Then we have that

$$\Delta \restriction \mathbb{F}(\{P, a\}) \Vdash P(a)$$
$$\Delta \restriction \mathbb{F}(\{P, b\}) \Vdash \neg P(b)$$

Consider $\theta = \Delta \restriction \mathbb{F}(\{P, a, b\})$ we may have $\theta \Vdash P(a)$ or $\theta \Vdash \neg P(a)$ or neither (but not both!). Similarly we have for the case of $P(b)$. We may now have that θ proves $P(a)$ and θ does not prove $\neg P(b)$ or θ proves $\neg P(a)$ and does not prove $P(b)$ or θ proves $P(b)$ and does not prove $\neg P(a)$ or θ proves $\neg P(b)$ and does not prove $P(a)$. In each of these cases we know how to revise.

If we still have that θ proves $P(a)$ and $\neg P(b)$ or θ proves $\neg P(a)$ and $P(b)$ or that θ proves nothing, then we revise arbitrarily.

Following the discussion of Remark 110, we can make a choice of whether to take $+P(a)$ or $\neg P(a)$ for our revised model. If $\Vdash$ does not tell us which one to take we can make an arbitrary choice. The algorithm is as follows:

1. If $\theta \Vdash P(a)$, then delete all occurrences of $\pm P(b)$ from $\varphi_{\mathbf{m}}$ to get $\varphi'_{\mathbf{m}}$. If $\theta \Vdash \neg P(b)$ then delete all occurrences of $\pm P(a)$ from $\varphi_{\mathbf{m}}$ to get $\varphi''_{\mathbf{m}}$. Otherwise delete $\pm P(b)$.

   Since we assume that $a = b$, we have that $\varphi'_{\mathbf{m}} = \varphi''_{\mathbf{m}}$ and this is our revised model. The revised theory $\mathbf{T}_{a/b}$ is the theory of this model.

Similar considerations will apply to $\mathbf{T}_{P/Q}$.

This completes our discussion of the non-monotonic identity merging method. This method is, however, not how the Talmud handles the case.

The above discussion of the obvious solution now has prepared us for the introduction of the Talmudic approach, as well as providing us with the means of comparison.

A theory can be revised by introducing new items of data which affect what it can prove. A theory can be revised also by cancelling or restricting the proof rules it can use. The latter method is used in resolving logical paradoxes. The data is fixed and leads to a paradox (inconsistency or unintuitive results). So one blocks some of the proofs and thus resolves the paradox.The Talmud revises by using a hierarchy of rules cancellations as we explain in the next section.

9.1 The Talmudic approach

Let us look at a well known example (x is universally quantified):

1. Bird(x) $\rightarrow$ Fly(x)
2. Penguin(x) $\rightarrow$ Bird(x)
3. Penguin(x) $\rightarrow \neg$ Fly (x)
4. Penguin (a)

We say that in viewing the above data, since Penguin is a more specific bird, then it wins and so we deduce $\neg$Fly(a).

Let us look now at the following data:

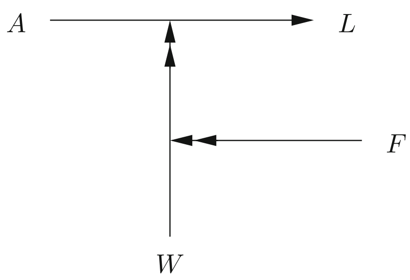


Figure 90.

5. Aeroplane 747 Flight BA101 $\rightarrow$ Land at Heathrow
   $A \rightarrow L$.
6. Aeroplane 747 Flight BA101 and Bad Weather $\rightarrow \neg$ Land at Heathrow
   $A \wedge W \rightarrow \neg L$
7. Aeroplane 747 Flight BA101 and Bad Weather and Short on Fuel $\rightarrow$ Land at Heathrow
   $A \wedge W \wedge F \rightarrow L$.

We may look at this again using the principle that the more specific assumptions (i.e. the antecedent of the rule contains more conjuncts than the other rule) win. So if we have only the information that an Aeroplane 747 Flight BA101 wants to land, we conclude that it can land. If we also add the conjunct that the weather is bad then it cannot land and if we even further add the conjunct that it is also short of fuel then it can land.

The Talmud looks at this differently as in Figure 90. W and F are meta-level principles. In the Figure ordinary arrow $\rightarrow$ means support and double arrow $\twoheadrightarrow$ means attack.

The basic principle is $A \rightarrow L$. The weather conditions involve a meta-level principle which cancel the arrow leading from A to L.[97]

The fuel shortage is involved in another meta-level principle which cancels the cancellation. So we are not dealing here with more specific knowledge but we are dealing with levels of meta-knowledge and a calculus of cancellations. The appropriate modelling of this is higher level attack and support (argumentation) networks.

So the Talmud uses a calculus of cancellations to resolve identity merging, as opposed to our previous proposal of non-monotonic support.

Let us give some examples.

EXAMPLE 112. This example is really from Talmudic logic, recast in everyday modern situation.

[97]Think of it as a rule of wisdom based on experience. "Just do not land in bad weather". Another such a rule is "If you are short of fuel land as soon as you can".

1. The story runs as follows:

   We have a duty to maintain our homes. We also have the instinct to save money. We believe in professional people doing jobs for us, but if we can do it properly ourselves, then we do it ourselves, and not call the expert and thus save money.

   So, if the kitchen sink is blocked, we do not call a plumber to do the job but do it ourselves and save money (a plumber home visit costs about $50 just to come, independent of the job he does).

   If the problem is more serious, say a blocked toilet, then better call a plumber and not take the risk of doing the job yourself. This case does need an expert!

   We can write these rules in non-monotonic logic as follows

   (a) x is blocked $\to$ repair x yourself

   (b) x is blocked $\wedge x$ is a serious job $\to$ get John the plumber to repair x

   (c) sink is blocked

   (d) toilet is blocked

   (e) repairing the toilet is a serious job, but not the sink.

   The problem with the above is that it implies that we call John the plumber and he repairs the toilet while we repair the sink. Common sense dictates that since the plumber is available he should repair the sink as well! We could add a new clause (f) to help:

   f. x is blocked $\wedge$ John the plumber repairs y $\to$ get John the plumber to repair x.

   Clause (f) says that if x is blocked and there is any y which John the plumber repairs y[98] then John the plumber to repair x. The format of clause (f) is not the usual monadic one, and does not make the information on x more specific. We can artificially fix this by adding a dummy universal predicate $U(x, y)$ which relates any two elements (something like $((x = y) \bigvee \neg(x = y)))$ and write (f*)

   f*. x is blocked $\bigwedge$ (John the plumber repairs $y \bigwedge U(x, y)$) $\to$ get John the plumber to repair x.

   Now (f*) is more specific than (a) on account of the additional predicate $V(x) =$ (John the plumber repairs $y \bigwedge U(x, y)$) This is clearly a fiddle and it departs from the intuitive understanding of what is going

[98]We tacitly assume here that they are all in the same, say, apartment building to be considered the same "call" by the plumber.

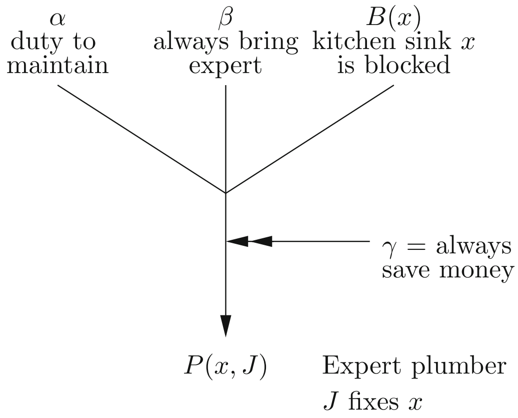


Figure 91.

on, which is clearly two meta-principles, namely save money but not at the expense of needed expertise!

Let us see how the Talmudic calculus of cancellations overcomes this problem.

Figures 91 and 92 describe these rules. The description is intuitive and not formal. The meaning of the nodes and arrows can be read intuitively from the figures.

The question arises what to do if both the sink and the toilet are blocked? If we just take the union of the two figures, (i.e. union of Figure 91 and Figure 92) i.e. update that the two plumbers a and b are equal, we will get that we call a plumber, the plumber does the toilet and at the same time we do the sink ourselves. It is more reasonable, however, since the plumber is already coming (and the \$50 call fee is to be paid anyway) to have the plumber do the sink as well.

Thus the "merging" of the two cases, i.e. merging of the two figures for the case that both the sink and the toilet are blocked is just a union of the graphs of the two figures. We will get Figure 93.

2. We now explain our notation.

   (a) $x, y, \ldots$ denote objects like x = kitchen sink, y = toilet.

   (b) B, P denote predicates which when applied to objects give states: $B(x)$ = kitchen sink is blocked, $B(y)$ = toilet is blocked. $P(x, z)$ = kitchen sink is repaired by plumber z, $P(y, z)$ = toilet is repaired by plumber z.

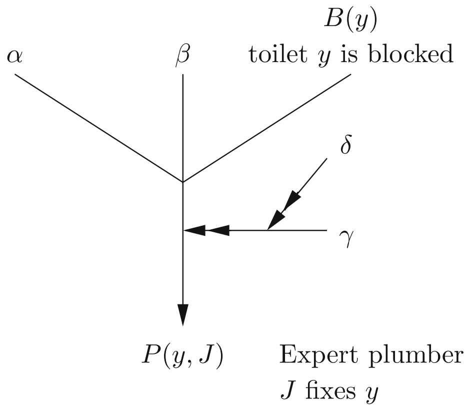


Figure 92.

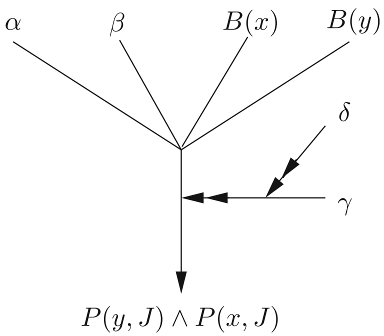


Figure 93.

$$B(x) \xrightarrow{\pi} P(x,a)$$

Figure 94.

$$\pi \text{ and } B(x) \to P(x,a)$$

Figure 95. Alternative notation to Figure 94

(c) α, β, γ are policies. For example:
α = policy to maintain your house
β = policy to always use experts
γ = policy to always save money
δ = policy to not take any risk for heavy maintenance jobs, if possible.

(d) A word about our notation: We denote the transition from one state to another by an arrow.

Figure 94 shows such notation. The π annotates the arrow. This means that because of policy π we take action and move from $B(x)$ to $P(x,a)$.

It may be that several policies come together and are involved in motivating some action, or it may be the case that some policies may cancel or overrule other policies. So we allow for alternative notation which we can use as well, when there are lots of policies to denote.

Figure 94 can be equivalently presented as Figure 95 or as Figure 96.

(e) Cancellation is done by double arrow.

Figure 97 shows some cancellations from some policies. It has no meaning, just a sample technical figure illustrating the notation.

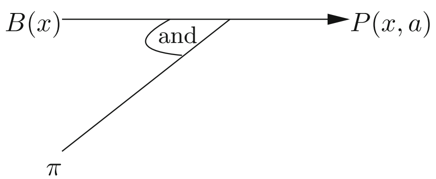


Figure 96. Alternative notation to Figure 94

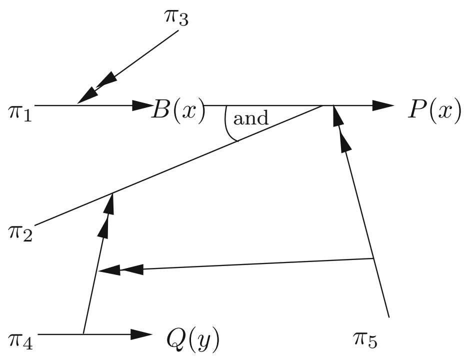


Figure 97.

i. π_1 and π_2 support together the move from $B(x)$ to $P(x,a)$.
ii. π_3 cancels the support of π_1 but allows the action to go forward on the basis of π_2.
iii. π_5 cancels the move to $P(x,a)$ no matter what, but also does not think that the support of π_4 to $Q(y)$ is a reason to cancel $\pi_2 \to P(x,a)$.

REMARK 113. The perceptive reader might think that the model of arrow cancellations as presented in Figures 91, 92 and 93 is nothing special and is just a notational variant of defeasible logic with specificity. Thus using the notation of Example 112 we can write a defeasible database Δ with the following universal formulas clauses, with w, z universal variables.

1. $B(w) \wedge \alpha \wedge \beta \wedge \gamma \to \neg P(w,z)$
2. $B(w) \wedge \alpha \wedge \beta \wedge \gamma \wedge \delta(w) \to P(w,z)$

If we instantiate (1) with $w = x, z = a$ and (2) with $w = y, z = b$ we get

1*. $B(x) \wedge \alpha \wedge \beta \wedge \gamma \to \neg P(x,a)$.
This is Figure 91 with $x =$ sink and $z =$ plumber a.

2*. $B(y) \wedge \alpha \wedge \beta \wedge \gamma \wedge \delta(y) \to P(y,b)$
This is Figure 92 with $y =$ toilet and $z =$ plumber b.

If we put (1*) and (2*) together in the same database and add the input $a = b = e$, then the database is consistent and the same plumber e will repair the toilet but not the sink. Defeasible logic based on specificity cannot tell us that because we have (2*) with $P(y,e)$ plumber e, we reverse and defeat (1*) and conclude $P(x,e)$ as well.

However, if we use the figures with the cancellation arrows, it is easier to model this feature. Figure 98 sums it all up. This is a predicate argumentation network involving joint attacks and higher order attacks, see [51; 53].

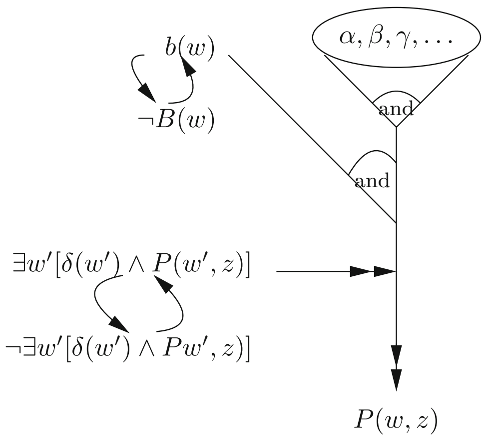


Figure 98.

9.2 Conclusion

We presented an outline in this section showing how Talmudic logic uses a calculus of cancellation to execute identity merging. In this conclusion section we want to impress upon the reader the schematic advantage of the calculus of cancellation.

Suppose we have two clauses

1. $\alpha \wedge A \to \exists x C(x)$
2. $\beta \wedge B \to \exists x \neg D(x)$.

We want to put (1) and (2) together in the same database and

(a) maintain consistency

(b) have the existential quantifiers pick up the same element.

The mechanisms we use are

(i) to take the specificity formulas out of the clauses and consider them as meta-principles, which are subject to being prioritised and apply to them the calculus of cancellations.

So we have

3. $\{\alpha, \beta\} : A \wedge B \to \exists x C(x) \wedge \exists x \neg D(x)$.

(ii) Convert

$$A \to \exists x C(x)$$
$$B \to \exists x \neg D(x)$$

into respective figures and turn (3) into

4. $\{\alpha, \beta\}$: union of Figures.

In the process of taking union of Figures we get that $\exists x$ chooses the same x.

REFERENCES

[1] J. L. Ackrill, (translator). *Aristotle's Categories and De Interpretatione*, Oxford University Press, 1963.

[2] M. Abraham, D. Gabbay and U. Schild. Paper 340 Kal-Vachomer in Hebrew, 112 pages. BDD Journal, Bar Ilan University.

[3] M. Abraham, D. Gabbay and U. Schild. Analysis of the Talmudic Argumentation A Fortiori Inference Rule (Kal-Vachomer) using Matrix Abduction. *Studia Logica*, 92:3, 281–364, 2009.

[4] M. Abraham, D. Gabbay, G. Hazut, Y. Maruvka and U. Schild. *The Textual Inference Rules Klal uPrat* (in Hebrew). College Publicatins, London, 2010. Hebrew paper 369. 200pp. 2009.

[5] M. Abraham, D. Gabbay and U. Schild. Obligations and prohibitions in Talmudic deontic logic. In *DEON 2010*, G. Governatori and G. Sartor, eds., pp. 166–178. LNAI 6181, Springer, 2010.

[6] M. Abraham, D. Gabbay and U. Schild. *Talmudic Deontic Logic*, College Publications, 2010.

[7] M. Abraham, D. Gabbay and U. Schild. *Future Determination of Entities in Talmudic Logic*. *Journal of Applied Logic*, 11(1), 63–90, 2013. http://dx.doi.org/10.1016/j.jal.2012.06.001

[8] M. Abraham, D. Gabbay and U. Schild. *Talmudic Temporal Logic*. College Publications, 2011, 600pp.

[9] M. Abraham, D. M. Gabbay, and U. Schild. The handling of loops in Talmudic logic with application to odd and even loops in argumentation. In *Proceedings of Howard 60*, D. Rydeheard, A. Vronkov and M. Korovina, eds., pp. 1–25, 2011.

[10] M. Abraham, G. Hazut, D. Gabbay, Maruvka and U. Schild. Logical Analysis of the Talmudic Rule of General and Specific (Klal-u-Prat). Special issue on Judaic Logic edited by A Schumann, *Journal of the History and Philosophy of Logic*, Vol 32 issue 1, pp 47–62, 2011.

[11] M. Abraham, I. Belfer, D. Gabbay and U.Schild. *Fuzzy Logic and Quantum States in Talmudic Reasoning*. College Publications, 2015.

[12] M. Abraham, D. Gabbay and U. Schild. *Principles of Talmudic Logic*. College Publications, 2013.

[13] M. Abraham, D. Gabbay and U. Schild. Obligations and prohibitions in Talmudic deontic logic. Expanded Journal version of [5]. *Journal of AI and Law*, 19:2, 117–148, 2011.

[14] M. Abraham, D. Gabbay and U. Schild. *Contrary to Time Conditionals in Talmudic Logic*. *Artificial Intelligence and Law*, 20:2, 146–179, 2012. DOI: 10.1007/s10506-9123-xOnlineFirst.

[15] G. Aucher, S. Barker, G. Boella, V. Genovese and L. van der Torre. Dynamics in Delegation and Revocation Schemes: A Logical Approach. In *Conference on Data and Applications Security and Privacy (DBSec'11)*, Richmond, Virginia USA. July 11-13, 2011

[16] O. Bandmann, M. Dam, and B. S. Firozabadi. Constrained Delegation. In *IEEE Symbposum on Security and Privacy*, pp. 131–140, 2002.

[17] E. Barka and R. Sandhu. Framework for Role-Based Delegation Models,. In *Proceedings of the 16th Annual Computer Security Applications Conference*, May 2000, pp. 168-176.

[18] S. Barker, G. Boella, D. Gabbay and V. Genovese. Reactive Kripke Models and Answer Set Programming Applications to Delegation and Revocation Schemes. *Journal of Logic and Computation*, 24:1, 89–116, 2014. doi: 10.1093/logcom/exs014.

[19] Steve Barker, Guido Boella, Dov M. Gabbay and Valerio Genovese. A Meta-model of Access Control in a Fibred Security Language, *Studia Logica*, 92(3), 437–477, 2009.

[20] P. Baroni and M. Giacomin. Solving semantic problems with odd-length cycles in argumentation. In *Proceedings of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2003)*, pp. 440–451. LNAI 2711, Springer-Verlag, Aalborg, Denmark, 2003.

[21] P. Baroni, M. Giacomi and G. Guida. SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168 (1-2):162–210, 2005.
[22] H. Barringer, D. Gabbay and J. Woods. Temporal dynamics of argumentation networks. In D. Hutter and W. Stephan, eds., *Mechanising Mathematical Reasoning*, pp. 59–98, LNCS 2605, Springer, 2005.
[23] Moritz Y. Becker, Jason F. Mackay and Blair Dillaway. Abductive Authorization Credential Gathering, *Studia Logica*, this issue.
[24] E. Bertino, S. Jajodia, and P. Samarati. Non-timestamped Authorization Model for Data Management Systems. In *3rd ACM Conference on Computer and Communications Security*, March 1996, pp. 169-178.
[25] P. Besnard and A. B. Hunter. *Elements of Argumentation*, 300pp. MIT Press, 2008.
[26] G. A. Bodanza and F. A. Tohmé. Two approaches to the problems of self-attacking arguments and general odd-length cycles of attack. *Journal of Applied Logic*, 7, 403–420, 2009.
[27] S. J. Brams, D. M. Kilgour, and W. S. Zwicker. The paradox of multiple elections. *Social Choice and Welfare*, 15:211–236, 1998.
[28] M. de Boer, D. Gabbay, X. Parent and M. Slavkova. Two dimensional standard deontic logic. *Synthese*, 187(2), 623–660, 2012. doi: 10.1007/s11229-010-9866-4.
[29] J. Broersen, A. Herzig and N. Troquard. From coalition logic to STIT. *Electronic Notes in Theoretical Computer Science*, 157, 23–35, 2006.
[30] A. Bochman. Collective argumentation and disjunctive logic programming. *J. Logic and Computation*, 13(3), 405–428, 2003.
[31] Martin Caminada. A Labelling Approach for Ideal and Stage Semantics. *Argument and Computation*, 2 (1): 1–21, 2011.
[32] M. Caminada. Preferred Semantics as Socratic Discussion, 2011.
[33] M. Caminada and D. M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.
[34] M. W. A. Caminada and Y. Wu. On the Limitations of Abstract Argumentation *BNAIC 2011*, Gent.
[35] M. W. A. Caminada and L. Amgoud. On the evaluation of argumentation formalisms, *Artificial Intelligence*, 171 (5–6), 286–310, 2007.
[36] Brian Chellas. *Modal Logic. An Introduction.* Cambridge University Press, 1980.
[37] Maria Luisa Dalla Chiara, Roberto Giuntini and Miklos Rédei. The History of Quantum Logic. In *Handbook of the History of Logic. Volume 8*, pp. 205–283, Dov M. Gabbay and John Woods, eds. Elsevier 2007.
[38] L. Cholvy and C. Garion. An Attempt to Adapt a Logic of Conditional Preferences for Reasoning with Contrary-To-Duties. *Fundamenta Informaticae*, 48, 2-3, 183–204, 2001.
[39] J. Carmo and A. J. I. Jones. Deontic Logic and Contrary-to-Duties. In D. M Gabbayand F. Guenthner, eds. *Handbook of Philosophical Logic*, vol 8, pp. 265-343, Springer, 2002.
[40] H. P. van Ditmarsch, W. van der Hoek and B. P. Kooi. *Dynamic Epistemic Logic*, Volume 337 of *Synthese Library*. Springer, 2007.
[41] Phan Minh Dung. An argumentation theoretic foundation for logic programming. *Journal of Logic Programming*, 22(2), 151–171, 1995.
[42] M. Eilon. *Jewish Law*, 4 volumes. Jewish Publication Society of America. Reprint edition, 1994.
[43] Final Proposed Draft Amendment on Certificate Extensions(v6). Generated from Collaborative ITU and ISO/IEC meeting on the Directory, April 1999. Orlando, Florida, USA.
[44] B. S. Firozabadi, M. Sergot, and O. Bandmann. Using Authority Certificates to Create Management Structures. In *Proceedings of Security Protocols, 9th International Workshop*, Cambridge, UK, pages 134–145. Springer Verlag, 2001.
[45] B. S. Firozabadi and M. Sergot. Revocation Schemes for Delegated Authorities. In *Proceedings of Policy 2002: IEEE 3rd International Workshop on Policies for Distributed Systems and Networks*. IEEE, June 2002.
[46] Menachem Fisch. *Rational Rabbis: Science and Talmudic Culture*. Indiana University Press, Bloomignton, 1997.

[47] R. Feynman, R. Leighton and M. Sands. *The Feynman Lectures on Physics, Vol III, Ch. 1.* Addison Wesley, Reading, 1965.

[48] D. M. Gabbay *et al. Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 1*, Oxford University Press, 1994.

[49] Dov Gabbay. The declarative past and imperative future: executable temporal logic for the interactive systems. In *Proc. Temporal Logic in Specification, LNCS 398*, pp. 409–448, Springer-Verlag, 1987.

[50] D. M. Gabbay *et al. Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 2*, Oxford University Press, 2000.

[51] D. Gabbay and M. Gabbay. Disjunctive attacks in argumentation networks, part 1. *Logic Journal of the IGPL*, 24:2, 186–218, 2016. doi: 10.1093/jigpal/jzv032 First published online: September 10, 2015

[52] D. Gabbay and J. Woods. Resource origins of non-monotonicity. *Studia Logica*, Vol. 88, 2008, pp. 85–112.

[53] D. Gabbay and A. Garcez. Logical modes of attack in argumentation networks. *Studia Logica*, 93(2-3), 199–230, 2009.

[54] D. M. Gabbay. *Labelled Deductive Systems*, OUP, 1996.

[55] D. M. Gabbay and J. Woods. *The Reach of Abduction*, Elsevier, 2005.

[56] T. K. Grabenhorst. *Das Argumentum A Fortiori.* Peter Lang, 1990.

[57] D. M. Gabbay and J. Woods. *Handbook of the History of Logic*, 12 Volumes, Elsevier, 2004–2011.

[58] D. Gabbay. Reactive Kripke semantics and contrary to duties, expanded version. *Journal of Applied Logic*, 11(1), 103–136, 2013.

[59] D. Gabbay and C. Strasser. Reactive Standard Deontic Logic. *Journal of Logic Computation*, 25:1, 1217–157, 2015. doi: 10.1093/logcom/exs043

[60] D. Basin, M. D'Agostino, D. Gabbay, S. Matthews and L. Vigano, eds. *Labelled Deduction*, Springer, 2000.

[61] D. Gabbay, G. Pigozzi, and J. Woods. Controlled Revision - An Algorithmic Approach for Belief Revision. *Journal of Logic and Computation*, 13(1):15-35, 2003.

[62] D. Gabbay, O. Rodrigues, and A. Russo. *Revision Acceptability and Context*, Springer, 2010.

[63] D. Gabbay. Reactive Kripke semantics and contrary to duty obligations. In *Deon 2008, Deontic Logic in Computer Science*, R. van der Meyden and L. van der Torre, eds., LNAI 5076, pp. 155–173, Springer, 2008.

[64]

[65] D. M. Gabbay. Dynamics of practical reasoning: A position paper. In *Advances of Modal Logic, Volume 2*, K. Segerberg, M. Zakhryaschev, M. de Rijke and H. Wansing, eds. pp. 197–242. CSLI Publications, CUP, 2001.

[66] D.M. Gabbay. Introduction to labelled deductive systems. In *Handbook of Philosophical Logic* Volume 17, second edition, 2013.

[67] D. Gabbay. Equational Approach to Argumentation Networks. *Argumentation and Computation*, 3(2-3), 87–142, 2012.

[68] D. Gabbay. Reactive Kripke Semantics and Arc Accessibility. In *Pillars of Computer Science: Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday*, Arnon Avron, Nachum Dershowitz, and Alexander Rabinovich, editors, Lecture Notes in Computer Science, vol. 4800, Springer-Verlag, Berlin, 2008 pp 292-341.

[69] S. A. Gaggl and S. Woltran. Strong Equivalence for Argumentation Semantics Based on Conflict-Free Sets, slides of lecture presented in ECSQARU 2011.

[70] G. Governatori and A. Rotolo. Changing legal systems: legal abrogations and annulments in defeasible logic. *Logic Journal of the IGPL*, 18: 157–194, 2010.

[71] M. Abraham, I. Belfer, D. Gabbay, and U. Schild. *Temporal Logic in the Talmud*, College Publications, 2011.

[72] D. Grossi and A. J. I. Jones Constitutive Norms and Counts-as Conditionals. In *Handbook of Deontic Logic*, edited by Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre. College Publications, 2013, pages 407-443.

[73] Å. Hagström, S. Jajodia, F. Parisi.Persicce, and D. Wijesekera. Revocation — a Classification. In *Proceeding of the 14th Computer Security Foundation Workshop*. IEEE press, 2001.
[74] J. Hansen. Conflicting imperatives and dyadic deontic logic. *Journal of Applied Logic*, 3, 484–511, 2005.
[75] S. Hartmann, G. Pigozzi and J. Sprenger. Reliable methods of judgement aggregation. Draft, 23 February 2009.
[76] A. Hasan. Analogical reasoning in Islamic jurisprudence, 1986. Republished Adam Publishers, 2007, 486 pp.
[77] A. Herzberg, Y. Mass, J. Mihaeli, D. Naor, and Y. Ravid. Access control meets public key infrastructure, or: Assigning roles to strangers. In *IEEE Symposium on Security and Privacy*, pages 2–14, 2000.
[78] E. Hirsch. Rashi's view of the Open Future: Indeterminateness and Bivalence. In *Oxford Studies in Metaphysics*, Oxford University Press, 2006.
[79] Jese Hughes. The muddy children. A logic for public announcement. Technical University of Eindhoven, February 10, 2007. http://phiwumbda.org/~jesse/slides/MuddlyChildren.handouts.pdf
[80] Louis Jacobs. *Studies in Talmudic Logic and Methodology*. London, Vallentine-Mitchell, 1061. Republished paperback, 2006.
[81] T. Jaeger, A. Edwards, and Xiaolan Zhang. Managing access control policies using access control spaces. In *Proceedings of the seventh ACM symposium on Access control models and technologies*, pages 3–12. ACM Press, 2002.
[82] A. J. I. Jones and M. Sergot. A formal characterization of institutionalised power. *Journal of the IGPL*, 3:427–443, 1996.
[83] M. H. Kamali. *Principles of Islamic Jurisprudence*, Islamic Text Society, 3rd Revised Eition, 2002, 550 pp.
[84] L. Kornhauser and L. Sager. Unpacking the court. *Yale Law Journal*, 96: 82–117, 1986.
[85] L. Kornhauser and L. Sager. The one and the many: adjudication in collegial courts. *California Law Review*, 81: 1–51, 1993.
[86] Arnold Kunst. An overlooked type of inference. *Bulletin of the School of Oriental and African Studies*, X, part 4, pp. 976–991, 1942.
[87] R. von Mises, Probability, Statistics and Truth [Dover, 1957], originally published in German [Springer, 1928].
[88] Jacob Neusner. The making of the mind of Judaism;the formative age. *Brown Judaic Studies*, vol 133, Scholars press, Atlanta, 1987.
[89] R. Nozick. *Anarchy, State and Utopia, 84–87*, 1974.
[90] R. Nozick. Coercion. In *Philosophy, Science, and Method: Essays in Hoor of Ernest Nagel*, S. Morgenbesser *et al.*, eds., pp. 440–447, 1969.
[91] G. Pigozzi and L. van der Torre. Premise independence in judgement aggregation. Dagstuhl Seminar 07531, 2007.
[92] H. Prakken and G. Sartor. Argument based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75, 1997.
[93] H. Prakken and M. Sergot. Contrary to duty obligations. *Studia Logica*, 57:1, 91–115, 1996.
[94] H. Prakken and M. Sergot. Dyadic deontic logic and contrary to duty obligations. In *Defeasible Deontic Logic*, D. Nute, ed., pp. 223-262. Synthese Library, Vol 263. Kluwer, 1997.
[95] H. Prakken. An abstract framework for argumentation with structured arguments *Argument & Computation*, 2010.
[96] E. Rissanen, B. S. Firozabadi, and M. Sergot. Towards a mechanism for discretionary overriding of access control, position paper. Presented at Security Protocols, *12th International Workshop*, Cambridge, UK, 2004.
[97] E. Rissanen, B. S. Firozabadi and M. Sergot. Discretionary Overriding of Access Control in the Privilege Calculus. *IFIP International Federation for Information Processing*, 2005, Volume 173, 219-232, 2005. DOI: 10.1007/0-387-24098-5_16
[98] U. J. Schild. Criminal Sentencing and Intelligent Decision Support, *Artificial Intelligence and Law*, vol. 6, 2-4, 1998.

[99] Erwin Schrödinger. *The Present Situation In Quantum Mechanics*, translator: John D. Trimmer. *Proceedings of the American Philosophical Society*, **124**:323-38. Reprint J. A. Wheeler and W. H. Zurek, eds., *Quantum Theory and Measurement*, Part 1, section I.11, Princeton University Press, New Jersey, 1983. See also article on Schroedinger's cat in *Wikipedia* http://en.wikipedia.org/wiki/Schroedinger's_cat

[100] Adolf Schwarz. *Der Hermeneutische Syllogismus in der Talmudischen*, Ltitteratur, Karlsruhe, 1901.

[101] J. Searle. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, 1969.

[102] J. Searle. *The Construction of Social Reality*. Free Press, 1995.

[103] V. L. S. Stebbing. *A Modern Introduction to Logic*, London, 1945.

[104] Andrew Schumann. Preface. In *History and Philosophy of Logic*, 32(1), pp. 1–8, 2011. doi = 10.1080/01445340.2010.506079, http://dx.doi.org/10.1080/01445340.2010.506079

[105] Andrew Schumann. Qal wa-omer and Theory of Massive-Parallel Proofs, *History and Philosophy of Logic*, 32(1), 71–83, 2011. doi = 10.1080/01445340.2010.506104, http://dx.doi.org/10.1080/01445340.2010.506104

[106] L. van der Torre and Y. H. Tan. Contrary-to-duty reasoning with preference-based dyadic obligations. *Annals of Mathematics and Artificial Intelligence*, 27:1-4, 49–78, 1999.

[107] Bart Verheij. A Labelling approach to to the computation of credulous acceptance in argumentation. In *Proceedings IJCAI'07. Proceedings of the 20th international joint conference on Artifical intelligence*, 623–628. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2007.

[108] T. Wakaki. Preference-based Argumentation Capturing Prioritized Logic Programming, *ArgMAS*, 2010.

INDEX

D. M. Gabbay, F. Guenthner (eds.), *Handbook of Philosophical Logic*,
Handbook of Philosophical Logic 18, https://doi.org/10.1007/978-3-319-97755-3