

A population choice experiment for testing axioms of stochastic discrete choice

William J. McCausland^{*1}, A.A.J. Marley² and Clinton P. Davis-Stober^{†3}

¹Département de sciences économiques and CIREQ, Université de Montréal, Canada

²Department of Psychology, University of Victoria, Canada

³Department of Psychological Sciences; MU Institute for Data Science and Informatics,
University of Missouri, U.S.A.

August 30, 2024

Abstract

We conducted an experiment suitable for testing axioms of stochastic discrete choice, including regularity, random utility and various forms of stochastic transitivity, for population level choice probabilities. Participants are representative of the Canadian adult population. The experiment features 32 different choice domains, including fine art, travel itineraries, and pizza toppings. For each domain, there is a set of five related objects, and for each of its 26 doubleton and larger subsets, we observe the choices of 40 participants. This allows us to measure population probabilities with some precision; since we do this for all non-degenerate menus of a given domain, we expose every implication of each of the axioms to possible falsification. Each participant faces exactly one menu from each domain, making this a strictly between-subject design; unlike similar studies with within-subject or mixed designs, the standard assumption that choices are independent, and on each menu, identically distributed, is innocuous. For each domain, we report Bayes factors giving evidence for or against each of the axioms.

Key words: Discrete choice, Choice axioms, Choice experiment, Context effects, Random utility, Bayesian inference

JEL codes: C11, C12, C15, C35, C91, D01, D90.

^{*}Corresponding author, william.j.mccausland@umontreal.ca, ORCID: 0000-0002-9896-235X. Data and code are publicly available at the Github repository <https://github.com/mccauslw/Population>.

[†]ORCID: 0000-0002-6836-6979

1 Introduction

Discrete choice models are used extensively in economics, marketing, psychology and other disciplines. Most empirical applications seek to measure how attributes of choice options affect choice probabilities in particular cases; for example, how commute times and other features affect the probabilities of choosing various transportation options, in Train (1978). Here, we contribute to a distinct and more abstract research agenda that collects evidence for or against various axioms on choice probabilities, without regard to object attributes. An example of such an axiom is the *regularity* condition, which states that for all menus A and B from a universe of objects such that $A \subseteq B$, the probability of choosing $x \in A$ when menu A is presented is no less than the probability of choosing x when B is presented. The other axioms we consider here are defined in Appendix A. One important objective of this research agenda is to find axioms that are empirically plausible across a wide variety of choice domains, which is important for theory refinement and development. To this end, we abstract away from object attributes and treat choice objects in a given universe as *a priori* identical. In different analyses, choice data observed over several trials may be the choices of a single decision maker or the choices of decision makers randomly selected from a population.

This paper advances this abstract choice research agenda by contributing new experimental data and some preliminary empirical analysis. The data serve the research agenda in several ways. First, the experiment includes a wide variety of choice domains. We observe participants choosing from menus from 32 different domains, including fine art, travel itineraries and pizza toppings. We can then see which choice patterns are more universal and which are more idiosyncratic. Second, for each domain, we observe choices from every possible non-degenerate menu. Each domain consists of five objects, and so there are 26 doubleton and larger menus that can be formed from the objects within. For each menu of each domain, we observe the choices of 40 different participants, allowing us to measure population choice probabilities with some precision. This allows us to expose all implications of axioms such as regularity to possible falsification. Third, the experiment features a strictly between-subject design. Each participant faces exactly one menu from each domain. This makes plausible the standard assumption that choices are independent, and on each menu identically distributed.

We provide an empirical analysis using a testing ground for discrete choice axioms developed in two previous papers. McCausland & Marley (2013) develops a non-parametric model for the system of choice probabilities on a domain. McCausland & Marley (2014) provides methods for posterior

simulation of parameters and choice probabilities and for testing axioms using Bayes factors. For each domain, we measure the evidence for or against each the seven discrete choice axioms defined in Appendix A.

1.1 Preliminaries

Let $T = \{x_1, \dots, x_n\}$ be a universe of choice objects. When faced with a non-empty menu $A \subseteq T$, an agent chooses a single object from A . The probability that the agent chooses $x \in A$ is denoted $P_A(x)$. A *random choice structure* (RCS) on T is the complete specification of the $P_A(x)$, $x \in A \subseteq T$, and is denoted P . Let Δ be the space of random choice structures, a Cartesian product of probability simplices.

We distinguish between two different interpretations of a RCS. An *individual* RCS governs the choices of a single individual; a *population* RCS, those of a random sample of individuals from a population. Note that if each individual in a population is governed by an individual RCS then the population RCS will be a convex combination of individual RCSs: each population choice probability mass function $P_A(\cdot)$, $A \subseteq T$, is a mixture—with random sampling from the population being the common mixing distribution—of the various individual probability mass functions $P_A(\cdot)$.

In situations where choice is repeated, we complete the probabilistic model by assuming that choices are statistically independent across trials, and that for each menu A , the same $P_A(\cdot)$ governs every choice from A . Following others, we refer to these two assumptions together as the iid assumption; note that “identically distributed” applies separately to each menu while “independent” applies globally.

There are good reasons to be sceptical of these assumptions in experiments featuring within-subject or mixed within/between designs; participants may be learning or their attention may be waning during the course of the experiment. While steps can be taken to attenuate these problems by limiting the number of repetitions of each menu and by using distractor trials to make it more difficult for participants to recognize choice objects, and while we can test the assumptions to some extent, we can never be sure that the assumptions hold. In a strictly between-subject design, however, with no communication between participants, the iid assumption is much more plausible.

1.2 Axioms of discrete stochastic choice

Various axioms, conditions, properties and hypotheses about probabilistic choice behaviour can be expressed as restrictions over the various choice probabilities of a RCS. Henceforth, we will

use the term axiom as a generic term to include all of these. Examples, all defined in Appendix A, include weak, moderate and strong stochastic transitivity, regularity, the triangle inequality, random utility (or the Block-Marschak inequality) and the multiplicative inequality. See Luce & Suppes (1965) for the various forms of stochastic transitivity, the triangle inequality and regularity; and McCausland & Marley (2013) for a summary of the logical relationships among all these axioms. Block & Marschak (1960) show that the Block-Marschak inequalities are necessary for random utility and Falmagne (1978) shows they are sufficient. The multiplicative inequality is due to Sattath & Tversky (1976), who show that it is a testable implication of the independent random utility model and also of the Elimination by Aspects (EBA) model described in Tversky (1972a) and Tversky (1972b). Tversky (1972a) shows that the Block-Marschak inequalities and moderate stochastic transitivity are also necessary conditions for, and thus testable implications of EBA.

Whether or not an axiom applies to a population RCS and whether or not it applies to individual RCSs are both of interest, but they are distinct questions. For example, weak stochastic transitivity (WST) is a compelling consistency principle for individual choice and empirical evidence against it is rare. However, de Condorcet's (1785) voting paradox casts doubt on the idea that population choice probabilities should satisfy WST.

Nonetheless, for some axioms, if the RCS for all individuals in a population satisfy the axiom, then the population RCS must as well. Expressed in contrapositive form to make plain the implications for testing, if the axiom is violated for a population, then it is violated by at least one individual. It is easy to see that the triangle inequality, regularity and random utility are examples of these axioms: each of these axioms consists of a set of simultaneous linear inequalities, and so the region where the axiom is satisfied is a convex set; the population RCS, being a convex combination of individual RCSs, must satisfy the axiom whenever all the individual ones do.

However, there are other axioms that can hold for every individual in a population but not the population itself. This is true of weak, moderate and strong stochastic transitivity, and the multiplicative inequality. We provide two counterexamples where a set of individual RCSs satisfying an axiom (or axioms) aggregate to population RCSs that do not. In both counterexamples, all individual RCSs are explicitly constructed as the choice probabilities resulting from an agent choosing the highest ranked object according to a random ranking. Thus, failure of aggregation is possible even if we require the individual RCSs to be well behaved in this sense.

Table 1 shows the two counterexamples. Lines 1, 2 and 3 give degenerate ranking and choice probabilities for three individuals (or equiprobable types of individuals) in a population. The choice

Line	p_i	π_{xyz}	π_{xzy}	π_{yxz}	π_{yzx}	π_{zxy}	π_{zyx}	$p(x, y)$	$p(y, z)$	$p(x, z)$	$P_T(x)$	$P_T(y)$
1	$\frac{1}{3}$	1						1	1	1	1	
2	$\frac{1}{3}$				1				1			1
3	$\frac{1}{3}$					1		1				
4		$\frac{1}{3}$			$\frac{1}{3}$	$\frac{1}{3}$		$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
5	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{2}$				$\frac{1}{2}$	1	1	$\frac{1}{2}$	$\frac{1}{2}$
6	$\frac{1}{2}$				$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}$			$\frac{1}{2}$
7		$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$

Table 1: Two aggregations of individual choice probabilities. Each line gives a random ranking (in columns π_{xyz} through π_{zyx} , where for example π_{xyz} is the probability of the ranking x, y, z) and the implied RCS (columns $p(x, y)$ through $P_T(y)$). The individual random rankings and RCS's in lines 1, 2 and 3 aggregate (with mixing probabilities p_i) to the population random ranking and RCS in line 4; likewise, lines 5 and 6 aggregate to line 7

behaviour of each individual satisfies weak, moderate and strong stochastic transitivity. The convex combination of the three individual RCSs, with equal weights, gives the population RCS of line 4. These probabilities violate even weak stochastic transitivity, since $p(x, y) = p(y, z) = 2/3$, but $p(x, z) = 1/3$. This first example is essentially de Condorcet's (1785) voting paradox. A second counterexample pertains to the multiplicative inequality. Lines 5 and 6 in Table 1 give choice probabilities for two individuals, each satisfying the multiplicative inequality. The equal-weighted convex combination of the two gives the population RCS of line 7, which violates the multiplicative inequality since $P_{\{x, y, z\}}(y) = 1/2 < p(y, x)p(y, z) = 3/4$.

While some axioms and conditions involve binary choice probabilities only, (e.g. weak, moderate and strong stochastic transitivity, the triangle inequality) other axioms involve multiple choice probabilities. Regularity, the multiplicative inequality and the Block-Marschak inequalities jointly constrain all choice probabilities on doubleton and larger menus. It is easy to see from the definitions of these axioms that in each case, if we are given an RCS satisfying the axiom and an arbitrary $A \subseteq T$ with at least two elements, we can always modify $P_A(\cdot)$ in such a way that the modified RCS violates the axiom. This illustrates the importance of collecting choice data for all doubleton and larger menus of a domain.

1.3 Context effects

One of the stated objectives of our experimental design is to include a wide variety of choice domains. To this end, we include choice domains designed to elicit anomalous behaviour, in the form of three

well documented “context effects”: the asymmetric dominance, similarity and compromise effects. In Appendix B, where we describe the choice domains in detail, we describe how some of the domains are designed to elicit these effects.

Here we describe versions of each of the three effects. Other versions of each of these context effects have been proposed, and they have different aggregation properties and different implications on whether systems of choice probabilities satisfy regularity, random utility (the Block Marschak conditions) and other conditions not considered here, like Krantz’s (1967) simple scalability.

The asymmetric dominance effect (sometimes known as the attraction effect) was introduced by Huber et al. (1982), and is defined up to a dominance relation on the universe of choice objects. In pertains to choice environments with three objects: a “target” x , a “competitor” y and a “decoy” z ; x “dominates” z , but y does not; neither x nor y dominates the other. The effect occurs when $P_{\{x,y,z\}}(x) > p(x, y)$, and constitutes evidence against regularity and random utility.

Tversky (1972b) introduces the similarity effect, defined up to the definition of a similarity relation on the choice objects under consideration. The similarity effect defined there occurs if there are x , y and z such that x and y are similar, x and z are not similar and y and z are not similar, and choice probabilities satisfy

$$p(x, z) > \frac{P_{\{x,y,z\}}(x)}{P_{\{x,y,z\}}(x) + P_{\{x,y,z\}}(z)} \quad \text{and} \quad p(y, z) > \frac{P_{\{x,y,z\}}(y)}{P_{\{x,y,z\}}(y) + P_{\{x,y,z\}}(z)}.$$

The inequality on the left holds if adding the object y , similar to the target object x , to the menu $\{x, z\}$ lowers the *relative* probability of choosing x , compared to z . The inequality on the right is similar, except that y is the target object and x is the similar object.

Simonson (1989) introduce a strong form of the compromise effect and Tversky & Simonson (1993) introduce a weaker version. They pertain to choice environments where objects can be described in terms of a set of real-valued attributes. An object y is *between* objects x and z , denoted $x|y|z$, if two conditions hold: first, for each attribute, y has a level of that attribute intermediate between the levels of x and z ; and second, that each of x and z have the highest level of at least one attribute. The weak effect is said to occur when $x|y|z$ and the following inequalities, from page 1183 in Tversky & Simonson (1993), hold:

$$p(y, x) < \frac{P_{\{x,y,z\}}(y)}{P_{\{x,y,z\}}(y) + P_{\{x,y,z\}}(x)} \quad \text{and} \quad p(y, z) < \frac{P_{\{x,y,z\}}(y)}{P_{\{x,y,z\}}(y) + P_{\{x,y,z\}}(z)}.$$

The inequality on the left holds if adding the object z to the menu $\{x, y\}$, making y a “compromise”

between x and z , raises the relative probability of choosing y , compared to x . The inequality on the right is similar, except that x is added to $\{y, z\}$, again making y a “compromise” between x and z .

2 Models

5 In this section we describe models for stochastic discrete choice data, models we will use to test various axioms. A *choice dataset* for a universe T consists of choices observed, over a number of trials, from various menus included in T . We denote a full dataset by \mathbf{X} when it is interpreted as a random variable, and by \mathbf{x} when interpreted as a possible value. A particular fixed choice of a RCS $P \in \Delta$, together with the iid assumption, fully describes a data generating process for \mathbf{X} , the
10 distribution $\mathbf{X}|P$.

A *Bayesian discrete choice model* consists of a prior distribution over Δ . Since the prior is a mixing distribution over choice probabilities, the mixture distribution also fully describes a data generating process for \mathbf{X} . For example, if we have a Bayesian discrete choice model M with prior density $f_M(P)$ on Δ , the mixture distribution is the marginal probability mass function
15 $f_M(\mathbf{x}) = \int_{\Delta} f(\mathbf{x}|P) f_M(P) dP$. Evaluation of $f_M(\mathbf{x})$ at the observed value \mathbf{x}_{obs} gives the *marginal likelihood* $f_M(\mathbf{x}_{\text{obs}})$ of the data, a number that measures the predictive performance of M for the observed \mathbf{x}_{obs} . Since it makes no reference to any unknown quantities, such as P , it can be interpreted as the out-of-sample prediction record of M for the observed data \mathbf{x}_{obs} .

To test axioms, we will compare *encompassing* models, Bayesian discrete choice models where
20 the prior density $f(P)$ has full support on Δ , with *constrained* models, where the axiom of interest limits the support of $f(P)$ to the region $\Lambda \subseteq \Delta$ where the axiom holds. We will use subscripts to denote regions associated with particular axioms, so that Λ_{SST} , for example, is the subset of Δ where strong stochastic transitivity holds. A constrained model is a Bayesian discrete model whose prior density over Δ is the truncation of the prior density in an encompassing model to Λ .
25 We will test axioms by computing Bayes factors—explained below—comparing constrained with encompassing models. We will provide prior sensitivity analysis by comparing these Bayes factors across a set of encompassing models.

Other studies have tested axioms by computing Bayes factors in favour of a constrained model against an encompassing model, including Cavagnaro & Davis-Stober (2014), Davis-Stober et al.
30 (2015), Myung et al. (2005) and Zwilling et al. (2011). These used simple non-hierarchical priors

on P where the $P_A(\cdot)$ are *a priori* stochastically independent and uniformly distributed. They analysed data featuring binary choice (or ternary choice where indifference between two objects is an option available to participants) and binary choice axioms. Prior distributions where the $P_A(\cdot)$ are *a priori* stochastically independent are feasible for testing weak and moderate stochastic
5 transitivity and the triangle inequality, because the implied prior probabilities of the axioms are high enough to measure using straightforward Monte Carlo. Strong stochastic transitivity, which is considerably more restrictive, as well as axioms that constrain all choice probabilities, not just the binaries, such as regularity, the multiplicative inequality and the Block-Marschak inequalities, tend to have much lower prior probability. For such axioms, it is important to introduce some stochastic
10 dependence among the various choice probability vectors $P_A(\cdot)$.

For the encompassing models, we use the prior distributions for P described in McCausland & Marley (2013), defined up to the values of two parameters, $\alpha \in (0, \infty)$ and $\lambda \in [0, 1]$. We will denote the prior density as $f(P|\alpha, \lambda)$. The parameter α describes how consistent choices are likely to be across trials. For each menu $A = \{a_1, \dots, a_{|A|}\}$, the marginal distribution of the
15 vector $(P_A(a_1), P_A(a_2), \dots, P_A(a_{|A|}))$ is symmetric Dirichlet with all $|A|$ shape parameters set to $\alpha/|A|$: for $\alpha \ll |A|$, draws from $P|\alpha, \lambda$ are likely to feature nearly deterministic choice, with choice probabilities close to zero or one; for $\alpha \gg |A|$, they are likely to be close to $1/|A|$; for $\alpha = |A|$, $P_A(\cdot)$ is uniformly distributed on the probability simplex. Figure 1 shows densities of binary choice probabilities, for five different values of α .

20 The parameter λ describes the degree of dependence of choice probabilities across menus. For $\lambda = 0$, the choice probability vectors $P_A(\cdot)$ are mutually independent, across $A \subseteq T$. For $\lambda = 1$, the random choice structure satisfies random utility with probability one. For $\lambda = 1$, and only this value, the prior does not have full support on Δ . Details on the prior distributions for P , including some of its desirable properties, are given in McCausland & Marley (2013).

25 It remains to specify the unknown parameters α and λ . One can choose fixed values, but a more flexible approach allows for learning about plausible values of α and λ through data. In this approach, one provides a prior distribution for α and λ , completing a hierarchical prior for P . Just as P can be marginalized out to obtain the marginal density $f_M(\mathbf{x})$ when using a non-hierarchical prior density $f_M(P)$ for P , we can marginalize out α , λ and P using a hierarchical prior consisting
30 of a prior density $f_M(\alpha, \lambda)$ for α and λ and the density $f(P|\alpha, \lambda)$ in McCausland & Marley (2013).

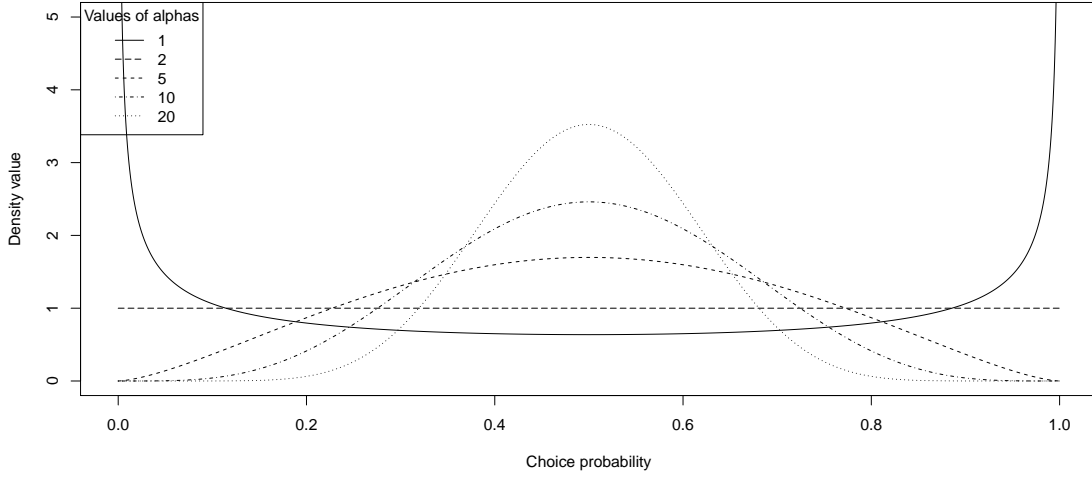


Figure 1: Density of binary choice probabilities for $\alpha = 1, 2, 5, 10, 20$

The result is the marginal density

$$f_M(\mathbf{x}) = \int f(\mathbf{x}|P)f(P|\alpha, \lambda)f_M(\alpha, \lambda) dP d\alpha d\lambda.$$

Again, the marginal likelihood $f_M(\mathbf{x}_{\text{obs}})$ is obtained by evaluating this at the observed choice dataset \mathbf{x}_{obs} .

Following McCausland & Marley (2014) and McCausland et al. (2020), we specify an encompassing model using a hierarchical prior distribution for P consisting of the prior density $f(P|\alpha, \lambda)$ and a prior distribution of (α, λ) where α and λ are *a priori* independent, α has a Gamma distribution with shape parameter $a_1 + a_2$ and rate parameter b , λ has a Beta distribution with shape parameters a_1 and a_2 . Like those papers, we use various choices of the hyper-parameter vector (a_1, a_2, b) , to illustrate the sensitivity of results to the choice of encompassing model. However, the particular choices of (a_1, a_2, b) we use here are different. In part, this is because the present paper is concerned with population, rather than individual choice. Anticipating some heterogeneity across individuals, we shift prior weight towards higher values of α , and therefore away from choice probabilities close to zero or one. Also, we learned from the analysis of these previous experiments that results were fairly insensitive to the prior for α , and that values of λ close to one were strongly favoured. Accordingly, we use the same prior for α for each encompassing model and put more prior weight on values of λ closer to one than in McCausland & Marley (2014) and McCausland

	a_1	a_2	b	$E[\alpha]$	$\text{Var}[\alpha]$	σ_α	$E[\lambda]$	$\text{Var}[\lambda]$	σ_λ
M_{ind}	0.00	3.00	3.00	9.00	27.00	5.196	0	0	0
M_1	2.00	1.00	3.00	9.00	27.00	5.196	0.667	0.0556	0.236
M_2	2.50	0.50	3.00	9.00	27.00	5.196	0.883	0.0347	0.186
M_3	2.75	0.25	3.00	9.00	27.00	5.196	0.917	0.0191	0.138

Table 2: Prior hyper-parameters and moments for various encompassing models

et al. (2020).

Table 2 shows the values of (a_1, a_2, b) for four encompassing models used in the empirical analysis. Model M_{ind} features a degenerate distribution for λ , with all prior mass at the value $\lambda = 0$, which makes the $P_A(\cdot)$ *a priori* independent. We will use this model only for testing the
5 less restrictive axioms (WST, MST, TI) as the prior probability of the other axioms under M_{ind} is too low to measure easily. The encompassing models M_1 , M_2 and M_3 feature dependence of $P_A(\cdot)$ across A and are suitable for testing strong stochastic transitivity and the multiple choice axioms.

Table 2 also indicates the prior mean, variance and standard deviation of α and λ for the four encompassing models. As we go from model M_1 to M_2 and then M_3 , the prior distribution of λ
10 concentrates more and more probability mass near $\lambda = 1$.

3 Inferential Methods

For any given choice dataset, our principal inferential objective is to test various axioms of discrete choice, using Bayes factors comparing models constrained by axioms to encompassing models. A secondary inferential objective, not pursued extensively here, is to describe the posterior distribution
15 of P and the parameters α and λ of the encompassing or the constrained models. Both inferential exercises are based on posterior simulation methods outlined in McCausland & Marley (2014).

The posterior distribution $\alpha, \lambda, P | \mathbf{x}_{\text{opt}}$, represents what we learn, relative to the prior distribution of α , λ and P , from the choice dataset \mathbf{x}_{opt} . The posterior distribution is analytically intractable and we use Markov chain Monte Carlo (MCMC) methods to obtain a large posterior
20 sample targeting the posterior distribution. The posterior sample does not consist of independent draws, but there are standard methods for quantifying simulation error; that is, the difference between posterior sample moments and their population counterparts. We use the overlapping batch means (OBM) method, whose properties are described in Flegal & Jones (2010), to compute the numerical standard error of posterior sample means.

We compute Bayes factors for all model comparisons. Suppose we are comparing any two Bayesian discrete choice models M and M' , and we observe a choice dataset \mathbf{x} . Then the posterior odds ratio giving the relative posterior probabilities of M and M' can be expressed as

$$\frac{\Pr[M|\mathbf{x}_{\text{opt}}]}{\Pr[M'|\mathbf{x}_{\text{opt}}]} = \frac{\Pr[M]}{\Pr[M']} \cdot \frac{f_M(\mathbf{x}_{\text{opt}})}{f_{M'}(\mathbf{x}_{\text{opt}})}$$

This result follows easily from Bayes' rule and does not depend on how many models are under consideration. The expression on the right hand side is the product of two factors. The first is the prior odds ratio and the second is a ratio of marginal likelihoods called the *Bayes factor*—see Berger (1985) or Bernardo & Smith (1994) for further explanation. The marginal likelihoods $f_M(\mathbf{x}_{\text{opt}})$ and $f_{M'}(\mathbf{x}_{\text{opt}})$ are the probabilities of observing the choice data \mathbf{x}_{opt} if the models M and M' (respectively) are true. The ratio measures the relative out-of-sample predictive performance of the two models, with a value greater than one favouring model M .

In the special case where the denominator model is an encompassing model M_e and the numerator model is the constrained model M_c obtained by truncating the density $f_{M_e}(P)$ associated with M_e to the region Λ defining an axiom, then the Bayes factor becomes

$$\frac{f_{M_c}(\mathbf{x}_{\text{opt}})}{f_{M_e}(\mathbf{x}_{\text{opt}})} = \frac{\Pr_{M_e}[\mathbf{X} = \mathbf{x}_{\text{opt}} | P \in \Lambda]}{\Pr_{M_e}[\mathbf{X} = \mathbf{x}_{\text{opt}}]} = \frac{\Pr_{M_e}[P \in \Lambda | \mathbf{X} = \mathbf{x}_{\text{opt}}]}{\Pr_{M_e}[P \in \Lambda]}. \quad (1)$$

The second equation of (1) follows from Bayes' rule; see Klugkist & Hoijtink. (2007) for details; the far right hand side of (1) is the ratio of posterior to prior probability of the axiom holding according to the encompassing model. A large numerator indicates an axiom that is highly probable in light of the data; a small denominator indicates that an axiom is restrictive. This ratio also suggests a practical way of computing approximations of Bayes factors: since the probability of an event is just the expectation of an indicator function for the event, we can approximate numerator and denominator using Monte Carlo methods.

We compute prior and posterior probabilities of axioms numerically using the simulation methods described in McCausland & Marley (2014). We compute numerical standard errors of prior and posterior probabilities using the OBM method and combine these standard errors to compute standard errors for Bayes factors using the delta method, introduced in Kelley (1928).

4 Experimental Design

The experiment consisted of an internet survey conducted between August 10, 2017 and August 31, 2017. A total of 1042 participants completed the survey. Supplemental materials related to the experiment, including all electronic files named in this section are available in the folder “Supplementary materials” of the Github repository at <https://github.com/mccauslw/Population>.

Participants were recruited by Survey Sampling International (SSI) from their Canadian panel, designed to be representative of the Canadian adult population. According to the SSI Global Panel Book 2017, `ssi-2017-panel-book.pdf`, the Canadian panel consists of 577,356 members. SSI screens potential members for reliability and requires that all participants be at least 18 years of age. They also carefully verify the identity of their participants for each survey. The Canadian panel is 42% male and 58% female and has the following age distribution: the percentage of members in the age range 18-24 is 36%; 25-34, 25%; 35-54, 26%; 55 and older, 13%. Compensation of participants includes cash and other rewards. A copy of the consent form that participants must complete before participating in the experiment is in the file `Consent_form.pdf`. A copy of the Certificate of Approval issued by Human Research Ethics Board at the University of Victoria is in the file `Certificate_of_approval.pdf`. A document describing in more detail the recruitment and compensation procedures followed by SSI is in the file `Recruitment_compensation_procedures.pdf`.

We chose not to impose any additional restrictions on our sample of participants. However, since invited participants must complete the survey to be included in the sample, there is some self-selection. We have full information on age, province and sex for all 1042 participants in our sample. There are 515 men (49.4%) and 527 women (50.6%). Age ranges from 18 to 88; the percentage of participants in the age bracket 18-24 is 11.4%; 25-34, 17.8%; 35-54, 38.3%, 55 and older, 32.5%. The number of participants in Alberta was 124; British Columbia, 136; Manitoba, 36; New Brunswick, 23; Newfoundland and Labrador, 16; Northwest Territories, 1; Nova Scotia, 29; Ontario, 391; Prince Edward Island, 4; Quebec, 251; Saskatchewan, 30; and Yukon, 1. Participants could take as long as they liked to complete the survey. The quartiles of total duration were 7.1, 10.0 and 13.9 minutes. Three participants took longer than 24 hours, and the shortest duration was 1.7 minutes. Participants were given the opportunity to leave written feedback on the experiment upon completion.

Programming and hosting of the survey were provided by the Institute for Choice at the University of South Australia.

Our experiment features $J = 32$ choice domains, including fine art, travel itineraries and pizza toppings. Each domain consists of a universe of five related choice objects. The names of the choice domains are shown in Table 3, together with the question (or an excerpt of the question, with excluded text indicated by an ellipsis) used to elicit a response from the participant. Appendix B
5 describes each domain in detail, including the question, the five choice objects and, where applicable, a source for the information used to construct the question. Figure 2 shows screenshots for the “Travel” and “Coffee” domains. There are $I = 2^5 - 5 - 1 = 26$ menus (binary and larger subsets of the five-element universe) for each domain, for a total of $J I = 910$ menus. The screenshot for the “Travel” domain in Figure 2 shows a menu with two of the five choice objects in that domain;
10 the “Coffee” screenshot, four of the five choice objects in the “Coffee” domain.

Regarding the assignment of participants to menus, our intended design was as follows. Within each domain, each of the I menus is presented exactly $N = 40$ times, once to each of N distinct participants. Each participant faces exactly one menu from each domain, for a total of J trials. Thus the total number of participants is $N I = 1040$, and the total number of trials is $N I J =$
15 $33,280$. For each domain j , the $N I$ participants are randomly partitioned into I groups of size N , with participants in each group seeing the same menu from that domain. Random partitions are uniformly distributed and mutually independent across domains. Elements of a menu are presented in random position (left, middle, right, for example) on the screen, independently across the N participants facing that menu. The J trials assigned to each participant are presented in
20 random order, independently across participants. Random partitions, positions and trial order are mutually independent.

In the event, two extra participants completed the experiment, and their choices were included in the analysis. For each domain, two menus were presented 41 times instead of 40 and the total number of trials is 33,344.



25 Participants choose exactly one object from each menu. For some domains, participants select radio buttons and in others, they select pictures, which are then identified by a border. Figure 2 shows one screenshot for each kind of domain. In each case, participants can modify their choice before they confirm it by clicking on a button labelled “>>”; once they click this button, they cannot go back and change their response. A progress bar indicates progress through the
30 experiment.

Data and code will be made available at the Github repository <https://github.com/mccauslw/Population>¹

¹The data and code are there now. Before eventual publication, we will be creating an R package for reproducing

Travel

Which one of the following travel destinations would you most like to visit?

Istanbul, TurkeyMarrakech, Morocco

>>

Coffee

You need to buy 16oz of ground coffee for a brunch with friends. Which one of the following ground coffees would you choose?

	Price (\$)	Fair Trade	Name: Description
<input type="radio"/>	18.71	Yes	Ethiopian Yirgacheffe: vibrant and intensely aromatic, fruity
<input type="radio"/>	9.99	No	Colombian Supremo: mellow cup, complex aromas and rich flavours
<input type="radio"/>	13.72	Yes	Colombian Organic: medium body, fragrant aroma and mild acidity
<input type="radio"/>	13.46	No	Sumatra Mandheling: exotic, earthy, bright with low acidity

>>

Figure 2: A screenshot for the “Travel” domain, in which participants select pictures, and a screenshot for the “Coffee” domain, in which participants select radio buttons. In the experiment and in electronic versions of this paper, these photographs are in colour.

Table 3: Choice domains of the experiment, with prompt question

Domain name	Question
Male stars	Which movie star would you choose to have lunch with?
Female stars	Which movie star would you choose to have lunch with?
Films	Judging from the following descriptions of films, which one of the films would you choose to see?
Star pairs	Knowing only who is starring, which one of these new films would you choose to see?
Pizzas	Which one of the following pizzas would you choose?
Juices	Which one of the following fresh juices would you choose?
Colours	Which one of the following colours do you like best?
Colour Pairs	Which one of these colour combinations do you like best?
Events	Which one of the following events do you think is most likely to happen in the next twenty years?
Radio formats	Suppose you were on a two hour road trip and you have a choice among radio stations with the following formats. Which one would you choose?
Music	Which one of the following musical artists do you like the best?
Aborig. art	Which one of the following examples of Australian aboriginal art do you like the best?
Impress. art	Which one of the following examples of Impressionist art do you like the best?
Sentences	Which one of the following sentences do you find the most grammatically acceptable?
Travel	Which one of the following travel destinations would you most like to visit?
Marijuana	Which one of the following marijuana policies would you choose?
Latitude	Which one of the following cities do you think is furthest north?
Dots	Which one of the following boxes do you think has the greatest number of points?
Triangles	Which one of the following triangles do you think has the greatest area?
Population	Which one of the following countries do you think has the largest population?
Area	Which one of the following countries do you think has the greatest surface area, including inland bodies of water?
Beer	... Given that you had to choose one brand to buy on this information alone, which one would you choose?
Cars	Which one of the following cars would you choose to drive, all other features begin equal? ...
Restaurants	Which one of the following restaurants would you choose for your next restaurant meal...?
Itineraries I	Which one of the following flight itineraries would you choose? ...
Payoffs	Which one of the following would you choose?
Phone plans	Of the following cell phone plans, which one would you choose? ...
Hotel rooms	... Which one of the following hotels would you choose, ...?
Itineraries II	Which one of the following flight itineraries would you choose? ...
Televisions	Which one of the following televisions would you choose to buy ...?
Coffee	Which one of the following ground coffees would you choose?
Charity	Suppose someone was donating a total of 100 dollars to a combination of charities, on your behalf. Which one of the following divisions of the 100 dollars would you choose?

the results of the paper. This will have better and more standard documentation and will facilitate the reproduction of the analyses on readers' own computers.

5 Results

We begin by providing some evidence that participants were engaged in the experiment. We then show some data summaries for one of the experimental domains, for illustrative purposes. The document `all_domains_data.pdf` in the folder “Supplementary materials” of the Github repository <https://github.com/mccauslw/Population> provides the same data summary for the other domains. We then report Bayes factors in favour of constrained models against encompassing models, for several choices of constraining axiom and encompassing model. Finally, we report posterior quantiles of α and λ for the various choice domains.

5.1 Engagement of participants

Evidence for a high level of engagement on the part of participants comes from written feedback, timing information, and the responses to questions where the objectively “correct” answer is known to most people. Most of the feedback left by participants was positive and many indicated a high level of interest. All feedback from participants is available in the document `feedback.pdf` in the folder “Supplementary materials” of the Github repository <https://github.com/mccauslw/Population>.

The median participant took 18.75s per domain; three quarters of them took longer than 13.31s per domain. The domain “Latitude”, where participants are prompted with the question “Which one of the following cities do you think is furthest north” is one of the domains whose elements have an objective order. In this domain, there are two pairs of cities where the objective order of latitude should be familiar to most people: one is “London, United Kingdom” and “Paris, France”; the other is “Vancouver, Canada” and “Seattle, United States”. We would expect most attentive participants to choose the city that is actually furthest north. Indeed, of the 40 participants offered the binary choice between London and Paris, 32 chose London and of the 40 participants offered the binary choice between Vancouver and Seattle, 36 chose Vancouver. Recall that the participants are residents of Canada.

5.2 Observed binary and ternary choice frequencies

Figure 3 illustrates all binary and ternary choice frequencies for the “Colours” domain, where the prompt is “Which one of the following colours do you like best?” and the universe of objects is $\{a, b, c, d, e\}$ where a is Red, b is Purple, c is Pink, d is Blue and e is Green. It also shows how these frequencies relate to regularity and the multiplicative inequality. Similar figures, as well as raw

data, for all 32 choice domains can be found in the document `all_domains_data.pdf` in the folder “Supplementary materials” of the Github repository <https://github.com/mccauslw/Population>.

The figure consists of 10 panels. Each panel corresponds to one of the ten tripton menus from that domain, and features an equilateral triangle, with vertices labelled with the three choice objects in that menu. Take the first panel as an example, where the corresponding menu is $A \equiv \{a, b, c\}$ and the vertices are labelled a (bottom left), b (top) and c (bottom right) accordingly. Each point in the interior or on the boundary of the triangle represents a triple $(P_A(a), P_A(b), P_A(c))$ of choice probabilities, in a Barycentric coordinate system. Thus, if triangle abc has unit height, the distance of a point to the right side of the triangle is $P_A(a)$, the distance to the base is $P_A(b)$ and the distance to the left side is $P_A(c)$. The point is also the convex combination of the vertices a , b and c —which have Barycentric coordinates $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, respectively—with weights $P_A(a)$, $P_A(b)$ and $P_A(c)$. The solid point in the interior of the triangle gives the triple $(\hat{P}_A(a), \hat{P}_A(b), \hat{P}_A(c))$ of choice frequencies observed for menu A of this domain, the sample counterpart of the ternary choice probability vector $(P_A(a), P_A(b), P_A(c))$.

The hollow dots on the left, right and bottom sides of the triangle represent binary choice frequencies $\hat{p}(a, b)$, $\hat{p}(b, c)$ and $\hat{p}(a, c)$, sample counterparts to the probabilities $p(a, b)$, $p(b, c)$ and $p(a, c)$, respectively. For example, the dot on the left side of the triangle is the convex combination of the vertices labelled a and b , with weights $\hat{p}(a, b)$ and $\hat{p}(b, a)$ respectively.

The downward pointing (blue in electronic versions of this paper) triangle in each figure gives necessary conditions on the ternary choice frequencies, given the three binary choice frequencies, for regularity to hold for choice frequencies on the sub-universe $T' \equiv \{a, b, c\}$: ternary probabilities within the blue triangle satisfy $P_A(a) \leq \min(\hat{p}(a, b), \hat{p}(a, c))$, $P_A(b) \leq \min(\hat{p}(b, a), \hat{p}(b, c))$ and $P_A(c) \leq \min(\hat{p}(c, a), \hat{p}(c, b))$. Similarly, the small upward pointing (red) triangle gives necessary conditions on the ternary choice frequencies, given the three binary choice frequencies, for the multiplicative inequality to hold for choice frequencies on T' : ternary probabilities in the red triangle satisfy $P_A(a) \geq \hat{p}(a, b)\hat{p}(a, c)$, $P_A(b) \geq \hat{p}(b, a)\hat{p}(b, c)$ and $P_A(c) \geq \hat{p}(c, a)\hat{p}(c, b)$.

5.3 Tests of axioms using Bayes factors

For each choice domain, we computed log Bayes factors in favour of several models, all relative to the encompassing model M_3 , which was favoured over M_1 and M_2 for 31 out of the 32 domains. We did this for the three encompassing models and for the constrained models defined by all possible pairs of encompassing model and axiom.

Figure 3: Binary and ternary choice frequencies for the Colours domain, where the prompt is “Which one of the following colours do you like best?” and the universe of objects is $\{a, b, c, d, e\}$, where a is Red, b is Purple, c is Pink, d is Blue and e is Green

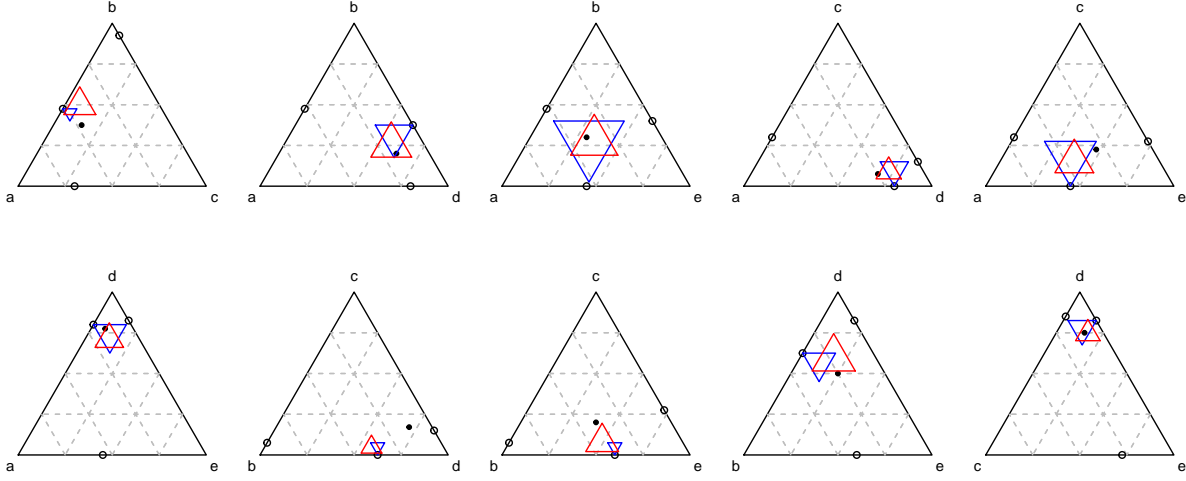


Figure 4 gives, for each domain, log Bayes factors for various combinations of binary choice axiom and encompassing model, relative to encompassing model M_3 . There is a panel for each domain, with three columns of points, corresponding to the three encompassing models M_1 , M_2 and M_3 . In each column, the circle gives the log Bayes factor of the corresponding encompassing model against M_3 ; in the third column, this value is always zero, since the log Bayes factor of any model against M_3 ; in the third column, this value is always zero, since the log Bayes factor of any model relative to itself is zero. The triangle, square and cross give log Bayes factors of the constrained model consisting of the restriction of the corresponding encompassing model to the regions Λ_{WST} , Λ_{MST} and Λ_{SST} , respectively. (Results for the remaining binary choice axiom, the triangle inequality, are reported below.) For the Charity choice domain, the posterior probability of SST is close to zero for all three encompassing models; while this makes numerical estimation of the log Bayes factor in favor of SST difficult, we can conclude that the evidence against SST is very strong² for this choice domain.

Across domains, the support for these binary choice axioms varies. With the exception of SST for a few domains (Male stars, Pizzas, Radio formats, Travel, Area and Charity) the evidence is moderate at best. For these few domains, the evidence is much stronger; interestingly, the evidence is sometimes in favour of SST and sometimes against. For most domains, the evidence for these

²Kass & Raftery (1995) classify log Bayes factors between 0 and 1 as “not worth more than a bare mention”, those between 1 and 3 as providing “positive evidence”, those between 3 and 5 as providing “strong evidence” and those above 5 as providing “very strong evidence”. We will adopt these categories, but refer to the first two as giving “weak” evidence and “moderate” evidence.

stochastic transitivity axioms is monotonic, in the sense that log Bayes factors are increasing or decreasing as we go from no axiom (none) to WST to MST to SST. A clear exception is the “Latitude” domain, where in decreasing order of Bayes factor, we have MST, then WST, then none then SST. In most domains, the order of log Bayes factors, and to some extent the differences, are robust to the encompassing model. For the exceptions, such as the Events domain, the differences are small and therefore there is little evidence supporting any one of these axioms over the others.

Figure 5 gives, for each choice domain, log Bayes factors for various combinations of multiple choice axiom and encompassing model, relative to encompassing model M_3 . It is organized in the same way as Figure 4. For the multiplicative inequality, the posterior probability is close to zero for the choice domains “Latitude”, “Itineraries I” and “Payoffs”, whatever the encompassing model, implying a strong rejection of that axiom. Evidence for the multiplicative inequality varies a lot across choice domain, from moderate evidence in favour to very strong evidence against. The evidence is fairly robust to the choice of encompassing model. For regularity and random utility, the evidence is weaker, but more consistent across choice domains. In most cases, the log Bayes factors either favour random utility and regularity over the encompassing models, or are very close to zero, representing little evidence either way. Across domains and encompassing models, the evidence in favour of random utility is very similar to the evidence in favour of regularity, which is a necessary condition for random utility. In most cases, the evidence in favour of regularity is greater than that in favour of random utility, indicating some evidence against the higher order features of random utility, but this evidence is quite weak. For individual domains, the evidence in favour of random utility and regularity are at best slight, but collectively, the evidence is much stronger: the log Bayes factor in favour of regularity holding jointly across domains is the sum over domains of the log Bayes factors in favour of regularity, and likewise for random utility. In both Figures 4 and 5, the variation in predictive performance of an axiom across underlying encompassing model is much less than the variation in that across the encompassing models themselves. This suggests similar implications for observables between imposing the axiom and increasing the prior dependence of $P_A(\cdot)$ across menus A , as measured by λ .

Table 4 shows Bayes factors in favour of WST, MST and TI against the encompassing model M_{ind} , where $\lambda = 0$ and choice probabilities $P_A(\cdot)$ are thus independent across menus. For comparison, it also shows Bayes factors in favour of these axioms against M_3 , the most favoured encompassing model for all domains except one. For encompassing model M_{ind} , the evidence favours WST, as it does for M_3 . However, the evidence is stronger, due to the relatively lower prior probability of

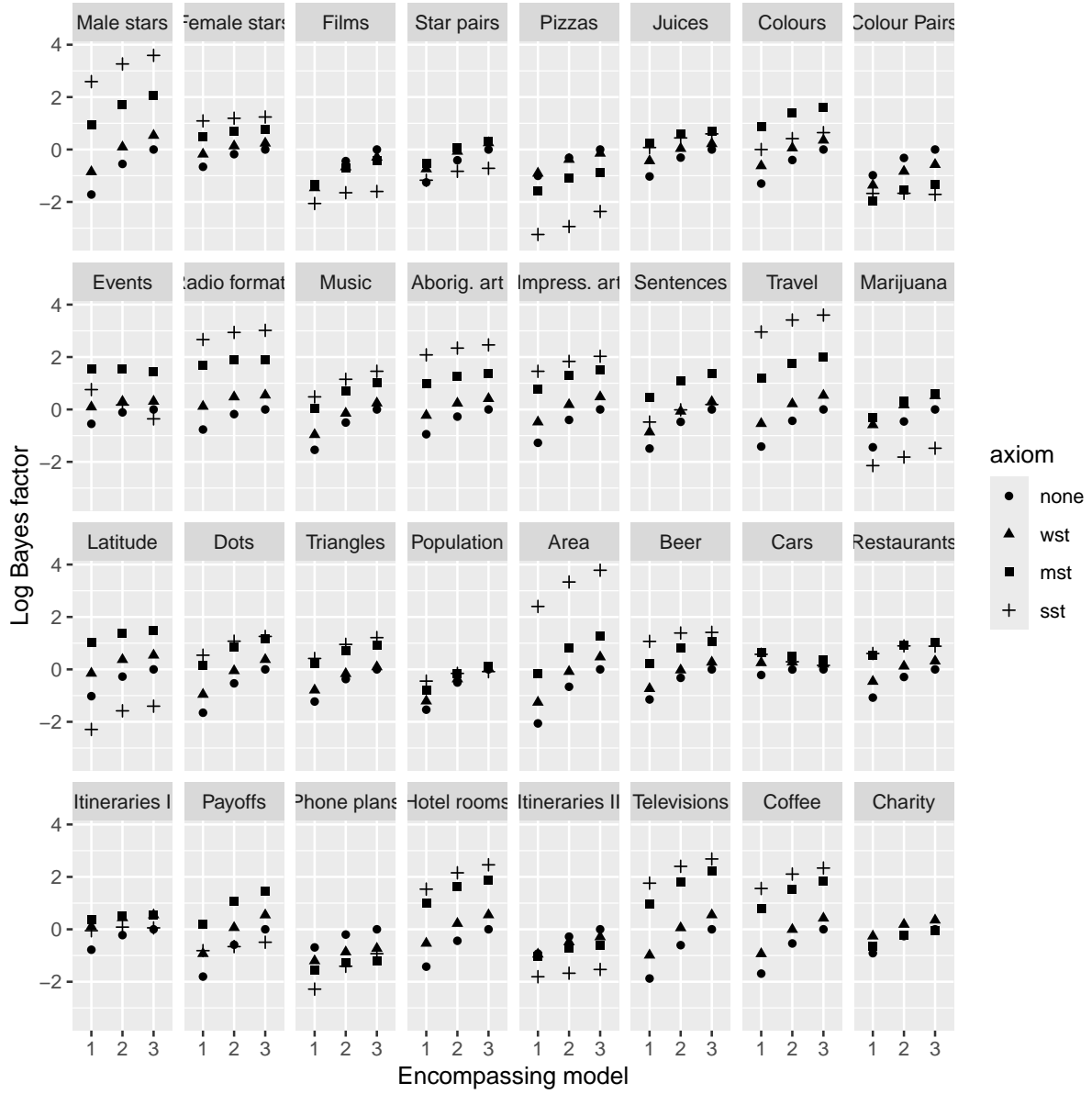


Figure 4: Log Bayes factors for 32 domains, by encompassing model and binary choice axiom. All log Bayes factors are relative to encompassing model M_3 . The “encompassing model” indices 1, 2 and 3 correspond to models M_1 , M_2 and M_3

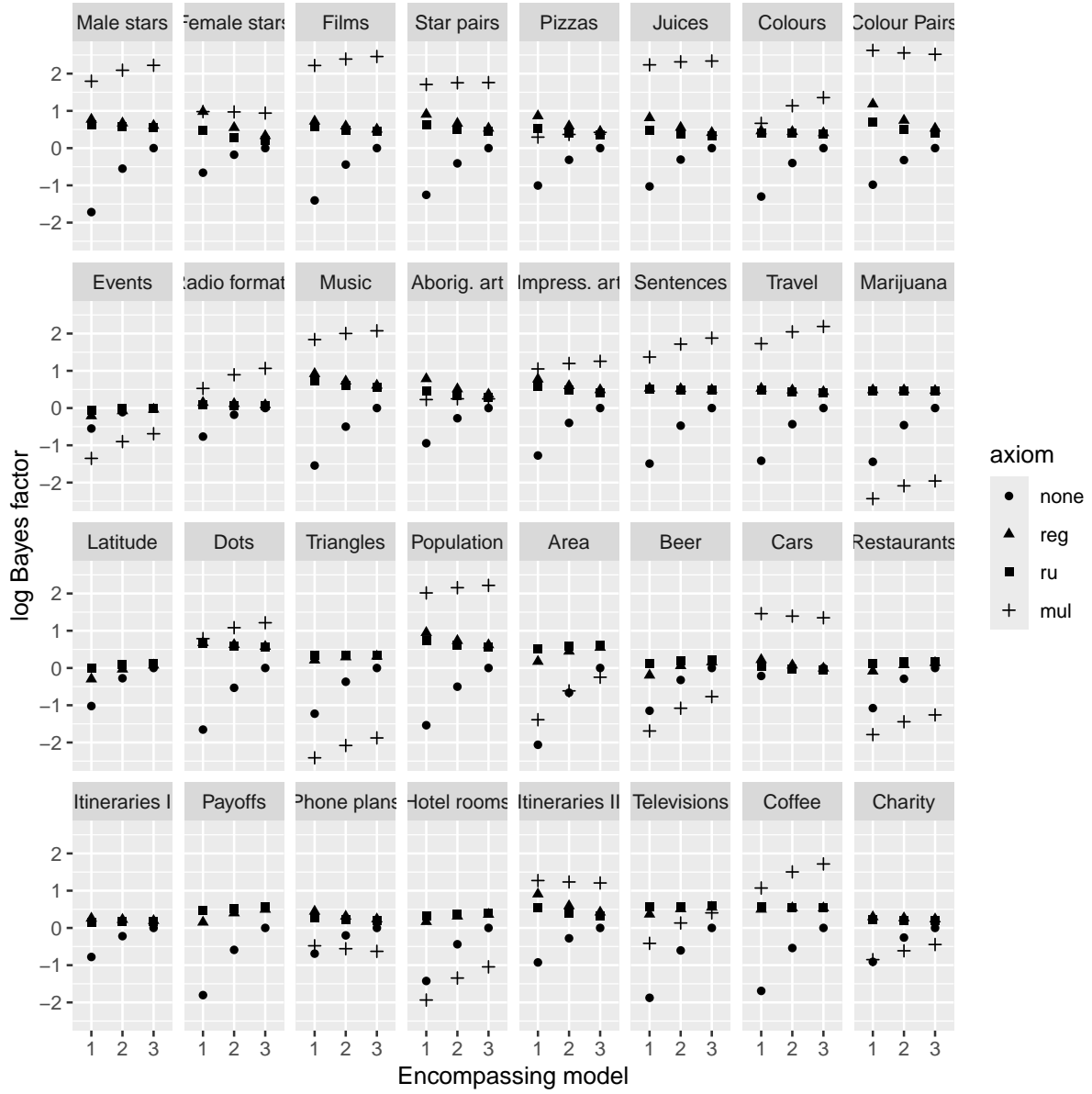


Figure 5: Log Bayes factors for 32 domains, by encompassing model and multiple choice axiom. All log Bayes factors are relative to encompassing model M_3 . The “encompassing model” indices 1, 2 and 3 correspond to models M_1 , M_2 and M_3

WST when the $P_A(\cdot)$ are independent. The two models mostly agree on whether MST is favoured or not. Again, the evidence is stronger for M_{ind} , and is either moderate or strong for most domains. Even in the four domains where the evidence is against MST, this evidence is weak.

For both encompassing models, the Bayes factor in favour of TI varies little across domain.
 5 This is because for most domains, the posterior probability of TI is close to one and so the log Bayes factor in favour of TI is close to its upper bound, which is equal to minus the logarithm of the prior probability of TI. The high prior probability of TI, especially for M_3 , limits the amount of evidence that the data can provide in favour of TI. Even the uniformly high posterior probability of TI across domains is less revealing than it may appear, in light of two observations. First, for
 10 any given choice domain, binary choice frequencies will tend to be more moderate for populations than for individuals, due to aggregation over heterogeneous individuals. Second, when binary choice probabilities are moderate, the triangle inequality is much less limiting. In fact, the triangle inequality holds for any RCS where all binary choice probabilities lie in the interval $[\frac{1}{3}, \frac{2}{3}]$. It is not very surprising then, that in this population choice experiment, binary choice frequencies tend to
 15 be more moderate than those in individual choice experiments, and that the posterior probability of the triangle inequality is very close to one.

5.4 Posterior distributions of α and λ

Table 5 shows numerical estimates of quantiles of α . The first row gives numerical estimates of prior quantiles for probabilities $p = 0.05$, $p = 0.5$ (i.e. the quantile is the median) and $p =$
 20 0.95 , together with their numerical standard errors. Subsequent rows give numerical estimates of posterior quantiles and their numerical standard errors, for the same probabilities, with a row for each domain. Plausible values of α are much higher than those reported in McCausland & Marley (2014) and McCausland et al. (2020), which analysed data from experiments with a within-subject design. Higher values of α imply more moderate choice probabilities, and so the results are
 25 unsurprising given the moderating effect of aggregation over heterogeneous individuals.

Table 6 shows numerical estimates of quantiles of λ . This parameter is of more theoretical interest because it governs the dependence of choice probabilities across menus and because of its connection with random utility. This table is similar to Table 5 except that the quantiles for $p = 0.05$, $p = 0.25$ and $p = 0.5$ are tabulated. The posterior quantiles of λ for $p = 0.75$ and
 30 $p = 0.95$ are very close to one for all domains and so the details are not very informative. This concentration of posterior probability very near $\lambda = 1$ is similar to what McCausland et al. (2020)

Domain	Model M_{ind}			Model M_3		
	WST	MST	TI	WST	MST	TI
Male stars	1.67	3.76	0.70	0.54	2.05	0.062
Female stars	1.03	1.75	0.71	0.24	0.75	0.062
Films	0.18	-0.43	0.50	-0.29	-0.41	0.061
Star pairs	0.54	0.42	0.70	0.25	0.32	0.062
Pizzas	0.21	-0.97	0.69	-0.14	-0.88	0.062
Juices	1.62	1.95	0.70	0.22	0.70	0.061
Colours	1.76	3.52	0.66	0.35	1.60	0.062
Colour Combinations	-0.20	-0.34	0.71	-0.58	-1.33	0.062
Events	1.68	3.94	0.63	0.31	1.43	0.060
Radio formats	2.05	3.94	0.68	0.55	1.91	0.061
Musical artists	1.67	2.30	0.70	0.24	1.02	0.062
Aboriginal art	1.71	3.18	0.71	0.41	1.38	0.062
Impressionist art	1.51	2.93	0.70	0.48	1.50	0.062
Sentences	1.72	3.47	0.70	0.29	1.38	0.062
Travel	2.07	4.09	0.70	0.54	1.99	0.062
Marijuana	1.79	1.81	0.42	0.52	0.61	0.060
Latitude	2.04	3.18	0.46	0.54	1.47	0.047
Dots	1.58	3.25	0.70	0.37	1.18	0.062
Triangles	1.39	2.09	0.68	0.09	0.91	0.061
Population	1.13	2.00	0.71	0.04	0.13	0.062
Surface area	1.88	4.20	0.70	0.47	1.28	0.062
Beer	0.56	1.61	0.42	0.27	1.06	0.054
Cars	1.13	2.06	0.70	0.20	0.36	0.061
Restaurants	1.53	2.95	0.64	0.32	1.02	0.060
Flight layovers	2.01	3.06	0.68	0.55	0.55	0.061
Future payments	1.85	3.52	0.70	0.55	1.47	0.061
Phone plans	0.19	-0.48	0.69	-0.73	-1.19	0.061
Hotel rooms	2.14	3.16	0.54	0.55	1.89	0.061
Two-flight itineraries	0.37	0.33	0.70	-0.29	-0.61	0.062
Televisions	2.07	4.77	0.69	0.55	2.21	0.062
Coffee	1.86	4.39	0.70	0.42	1.84	0.062
Charity	1.39	0.83	0.55	0.35	-0.06	0.058

Table 4: Log Bayes factors in favour of binary choice axioms, relative to encompassing models M_b and M_3 , by domain. The largest numerical standard errors for the six columns are 0.011, 0.090, 0.004, 0.014, 0.043 and 0.001

Domain	p=0.05		p=0.5		p=0.95	
	quant	nse	quant	nse	quant	nse
Prior	2.45	0.005	8.02	0.007	18.84	0.020
Male stars	6.50	0.137	12.38	0.124	21.11	0.155
Female stars	8.53	0.161	16.02	0.127	27.27	0.166
Films	6.67	0.133	12.64	0.128	21.78	0.175
Star pairs	7.20	0.145	14.26	0.149	25.44	0.201
Pizzas	7.71	0.130	14.30	0.124	24.22	0.170
Juices	7.47	0.120	14.24	0.127	24.49	0.177
Colours	5.66	0.077	10.12	0.096	16.85	0.144
Colour Combinations	9.06	0.117	17.42	0.130	30.49	0.181
Events	4.48	0.070	8.23	0.085	13.78	0.129
Radio formats	5.70	0.108	10.55	0.114	17.80	0.174
Musical artists	7.82	0.114	14.43	0.117	24.31	0.158
Aboriginal art	7.64	0.108	14.24	0.130	24.26	0.172
Impressionist art	7.23	0.184	13.45	0.123	22.51	0.176
Sentences	6.55	0.118	11.84	0.112	19.77	0.163
Travel	6.08	0.219	11.53	0.113	19.07	0.153
Marijuana	4.59	0.110	9.07	0.124	15.77	0.178
Latitude	2.93	0.055	5.74	0.084	10.15	0.140
Dots	6.71	0.086	12.00	0.103	19.80	0.151
Triangles	4.92	0.074	8.99	0.103	15.49	0.179
Population	7.97	0.121	14.69	0.115	24.96	0.174
Surface area	3.08	0.044	5.62	0.061	9.39	0.096
Beer	3.90	0.076	7.25	0.090	12.45	0.146
Cars	7.01	0.134	13.36	0.120	23.16	0.178
Restaurants	4.47	0.080	8.29	0.099	14.20	0.139
Flight layovers	4.86	0.136	9.81	0.129	16.91	0.177
Future payments	3.09	0.077	6.06	0.075	10.36	0.116
Phone plans	6.32	0.116	12.48	0.140	22.76	0.184
Hotel rooms	3.76	0.077	7.24	0.103	12.65	0.154
Two-flight itineraries	8.30	0.115	15.46	0.122	26.60	0.172
Televisions	4.52	0.096	8.28	0.091	13.84	0.134
Coffee	5.72	0.080	10.18	0.098	16.80	0.144
Charity	4.89	0.104	9.83	0.138	17.78	0.201

Table 5: Numerical estimates of prior quantiles of α , and of posterior quantiles of α by choice domain, with their numerical standard errors, for the encompassing model M_3 . Quantiles are for probabilities 0.05, 0.50 and 0.95

found for a large majority of the 140 participants in the within-subject experiment described there. However, while there were a few exceptions in that previous experiment, there are no domains in the present experiment that are exceptional in this regard. This gives strong evidence for a kind of posterior dependence of the $P_A(\cdot)$ across menus A of a kind that is related to random utility.

5 In some sense, putting higher prior probability mass near $\lambda = 1$ on the one hand, and imposing random utility on the other, have similar implications for observables: Figure 5 above shows that for every domain, the Bayes factors in favour of random utility (the three black squares in each panel) vary much less across the three underlying encompassing models than the Bayes factors in favour of the encompassing models themselves (the black circles).

10 6 Conclusions

We have collected new experimental data for population choice probabilities. Several features of the experimental design make the data suitable for testing axioms of discrete probabilistic choice for population level choice.

First, it includes a wide variety of choice domains. Some choice domains involve subjective
 15 aesthetic judgements, on such matters as food and drink, art, music, politics and the grammaticality of sentences. Other choice domains resemble choice domains that naturally occur in markets, such as those for mobile phone plans, hotel rooms, beer, cars and flights. Still others involve options having an objective rank order. Some choice domains have options that are distinguished only by the levels of numerical attributes, others do not. Some domains are designed to elicit context
 20 effects, others are not. This variety helps us and others to discover properties of random choice structures that hold widely. An unintended but welcome additional consequence was the high level of engagement of the participants.

Second, for each choice domain, we observed repeated choices from all 26 doubleton and larger subsets of a universe of five objects, allowing us to test any axiom of discrete probabilistic choice by
 25 exposing every implication of it to possible falsification. Third, by arranging that every participant saw exactly one menu from each choice domain, we obtain data on population choices that are plausibly independent as well as identically distributed, menu by menu.

The experimental data we have collected complements the models introduced in McCausland & Marley (2013) and the inferential methods described in McCausland & Marley (2014). Together,
 30 the models, methods and data serve as a testing ground for the axioms we have considered in this

Domain	p=0.05		p=0.25		p=0.50	
	quant	nse	quant	nse	quant	nse
Prior	0.606	0.0009	0.897	0.0003	0.9819	0.00008
Male stars	0.934	0.0012	0.986	0.0003	0.9977	0.00004
Female stars	0.817	0.0036	0.950	0.0012	0.9907	0.00023
Films	0.908	0.0019	0.980	0.0004	0.9966	0.00007
Star pairs	0.891	0.0025	0.978	0.0005	0.9963	0.00007
Pizzas	0.854	0.0031	0.967	0.0008	0.9945	0.00012
Juices	0.866	0.0025	0.968	0.0007	0.9942	0.00012
Colours	0.897	0.0022	0.976	0.0006	0.9957	0.00009
Colour Combinations	0.839	0.0038	0.966	0.0008	0.9943	0.00012
Events	0.800	0.0055	0.932	0.0026	0.9831	0.00089
Radio formats	0.835	0.0038	0.947	0.0014	0.9873	0.00045
Musical artists	0.915	0.0015	0.983	0.0003	0.9971	0.00005
Aboriginal art	0.841	0.0031	0.959	0.0009	0.9923	0.00017
Impressionist art	0.899	0.0021	0.979	0.0005	0.9964	0.00007
Sentences	0.917	0.0014	0.982	0.0004	0.9969	0.00006
Travel	0.916	0.0016	0.980	0.0005	0.9966	0.00010
Marijuana	0.925	0.0019	0.984	0.0004	0.9973	0.00006
Latitude	0.873	0.0038	0.967	0.0013	0.9937	0.00031
Dots	0.930	0.0015	0.986	0.0003	0.9976	0.00004
Triangles	0.900	0.0023	0.976	0.0007	0.9957	0.00011
Population	0.921	0.0012	0.983	0.0003	0.9972	0.00004
Surface area	0.957	0.0010	0.991	0.0002	0.9985	0.00003
Beer	0.890	0.0027	0.969	0.0010	0.9940	0.00023
Cars	0.732	0.0057	0.909	0.0029	0.9799	0.00083
Restaurants	0.883	0.0029	0.968	0.0010	0.9936	0.00022
Flight layovers	0.821	0.0052	0.951	0.0019	0.9906	0.00039
Future payments	0.943	0.0016	0.989	0.0003	0.9981	0.00004
Phone plans	0.798	0.0059	0.947	0.0017	0.9899	0.00032
Hotel rooms	0.919	0.0020	0.981	0.0005	0.9966	0.00009
Two-flight itineraries	0.833	0.0033	0.960	0.0009	0.9930	0.00016
Televisions	0.950	0.0010	0.990	0.0002	0.9983	0.00003
Coffee	0.939	0.0012	0.987	0.0003	0.9978	0.00004
Charity	0.852	0.0033	0.959	0.0013	0.9920	0.00028

Table 6: Numerical estimates of prior quantiles of λ , and of posterior quantiles of λ by choice domain, with their numerical standard errors, for the encompassing model M_3 . Quantiles are for probabilities 0.05, 0.25 and 0.50

paper and any other axioms that one may wish to investigate.

We have known since the introduction of de Condorcet’s (1785) celebrated paradox not to expect population choice probabilities to be even weakly stochastically transitive. However, in most cases there is weak evidence in favour of weak stochastic transitivity and in the few remaining cases, the evidence against is slight. Evidence for or against moderate and especially strong stochastic transitivity tends to be stronger and more variable. Clearly, these are not promising candidates for axioms that hold widely, but the fact that strong stochastic transitivity is strongly favoured for some domains is intriguing and raises the question of what features these domains might have in common. The triangle inequality has very high posterior probability for most of these domains, but this is largely to do with the anodyne fact that observed binary choice proportions are quite moderate, as we might expect them to be for population level data due to the moderating effect of aggregation. We conclude that even if the triangle inequality makes sense for population choice probabilities, it will not have much empirical traction for many choice environments.

Evidence for regularity and random utility is fairly consistent across choice domains. Together with the triangle inequality, which is a necessary condition for both, these are the axioms considered here that correspond to convex regions of the space of RCSs, and which therefore aggregate across individuals. While the evidence in favour of these for any one choice domain is relatively weak, the collective evidence is quite favourable. It is important to note too, that the encompassing models used for computing Bayes factors already feature a great deal of prior dependence among choice probability vectors $P_A(\cdot)$ (as measured by λ) similar to the kind of dependence associated with random utility, and that encompassing models with a higher prior mean of λ are favoured. Evidence for the multiplicative inequality, like that for strong stochastic transitivity, varies considerably across domains. This is not a promising candidate for an axiom that holds widely either, but again, the fact that it is favoured in many domains, although not as strongly as strong stochastic transitivity, raises the question of what features these domains might have in common.

In future work we wish to investigate Sattath & Tversky’s (1976) multiplicative inequality in more detail, using a variety of data sources, complementing some theoretical work on that condition in McCausland & Marley (2024). Using the data from the experiment described in the present paper, we hope to explore the occurrence (or not) of context effects in those domains that are designed to elicit them, and also to see if similar patterns of probabilities occur in domains where choice objects do not have obvious numerical attributes. In ongoing work, McCausland (unpublished) computes marginal likelihoods for random preference models corresponding to $\lambda = 1$

in the model of McCausland & Marley (2013). This will allow for cleaner tests of random preference and random utility models, against a wider set of alternative models.

A Definitions of axioms and basic results

A random choice structure satisfies

- 5 **WST** *weak stochastic transitivity* if and only if for all distinct x, y and z ,

$$p(x, y) \geq \frac{1}{2} \text{ and } p(y, z) \geq \frac{1}{2} \Rightarrow p(x, z) \geq \frac{1}{2},$$

MST *moderate stochastic transitivity* if and only if for all distinct x, y and z ,

$$p(x, y) \geq \frac{1}{2} \text{ and } p(y, z) \geq \frac{1}{2} \Rightarrow p(x, z) \geq \min[p(x, y), p(y, z)],$$

SST *strong stochastic transitivity* if and only if for all distinct x, y and z ,

$$p(x, y) \geq \frac{1}{2} \text{ and } p(y, z) \geq \frac{1}{2} \Rightarrow p(x, z) \geq \max[p(x, y), p(y, z)],$$

TI the *triangle inequality* if and only if for all distinct x, y and z ,

$$p(x, y) + p(y, z) + p(z, x) \geq 1,$$

Reg *regularity* if and only if for all $x \in A \subseteq B \subseteq T$,

$$P_A(x) \geq P_B(x),$$

- 10 **MI** the *multiplicative inequality* if and only if for all $x \in A, B \subseteq T$,

$$P_{A \cup B} \geq P_A(x)P_B(x),$$

RU *random utility* (or the the *Block-Marschak inequality*) if and only if for all non-empty $A \subseteq T$ and all $x \in A$,

$$\sum_{B: A \subseteq B \subseteq T} (-1)^{|B \setminus A|} P_B(x) \geq 0. \quad (2)$$

B Domains

Here, we describe a set of J choice domains.

We classify choice domains into four categories. In the first category, objects have no numerical attributes. In the second category, objects have a single attribute, whose level is not explicitly
5 given. In the third category, objects have two numerical attributes and the experimental design is intended to elicit one or more context effects. In the fourth category, objects have many attributes, and they are chosen to resemble objects in discrete choice experiments used in Marketing, although with fewer attributes.

The section titles, “Male stars” for example, give the names of the domains for the purpose of
10 reporting results. The specification of a domain is shown in a grey box and consists of a question and five choice objects, or responses. In the actual experiment, for a given domain, all participants see the same question (such as “Which movie star would you choose to have lunch with?”) but not the same set of responses; different participants will see from two to five of the possible responses, in random order, as described in Section 4. For the purpose of reporting results, the choice objects in
15 each domain are labelled a , b , c , d and e , in the order presented in the grey boxes. The participants do not see these labels, and accordingly, we do not use them within the boxes.

We adopt the convention that whenever there is an established order for the items in a list, we order them in decreasing order. Again, this is only for the purposes of reporting results; each participant sees the items in a menu in random order.

20 B.1 No numerical attributes

B.1.1 Male stars

The source for this domain is the website `ranker.com`, accessed June 4, 2017. The list is “The best actors working today”. The choices are the top five actors in that list, in order.

Which movie star would you choose to have lunch with?

- Tom Hanks
- Kevin Spacey
- Morgan Freeman
- Leonardo DiCaprio
- Christian Bale

B.1.2 Female stars

The source for this domain is the website ranker.com, accessed June 4, 2017. The list is “The best American actresses working today”. These are the top five actors in that list, in order. Jodie Foster’s name was misspelled in the experiment, as two participants noted in the comments.

Which movie star would you choose to have lunch with?

- Meryl Streep
- Jody Foster
- Kathy Bates
- Amy Adams
- Julianne Moore

5

B.1.3 Films

The source for this domain is the IMDb list “Most Popular Feature Films Released 1990 to 1999”. The decade was chosen so that the films would not be easily recognizable by most respondents.

Judging from the following descriptions of films, which one of the films would you choose to see?

- Two imprisoned men bond over a number of years, finding solace and eventual redemption through acts of common decency.
- Mathilda, a 12-year-old girl, is reluctantly taken in by Léon, a professional assassin, after her family is murdered. Léon and Mathilda form an unusual relationship, as she becomes his protégé and learns the assassin's trade.
- The lives of two mob hit men, a boxer, a gangster's wife, and a pair of diner bandits intertwine in four tales of violence and redemption.
- A sexually frustrated suburban father has a mid-life crisis after becoming infatuated with his daughter's best friend.
- Identical twins, separated at birth and each raised by one of their biological parents, discover each other for the first time at summer camp and make a plan to bring their wayward parents back together.

B.1.4 Star pairs

Here, choice objects are pairs of movie stars from a set of four movie stars: Tom Hanks, Scarlett Johansson, Brad Pitt and Angelina Jolie. The only missing pair is Brad Pitt and Angelina Jolie.

- 5 One possible measure of similarity is the number of actors in common between two pairs, with values 0 and 1. There are nine doubleton choice sets (i.e. pairs of actor pairs) with one star in common and one ($\{c, d\}$) without any stars in common. Thus there are three triples ($\{a, c, d\}$, $\{b, c, d\}$, $\{c, d, e\}$) where one might expect a similarity effect.

In this example, respondents' preferences may depend not only on their liking of particular
10 actors but also on complementarities between actors.

Knowing only who is starring, which one of these new films would you choose to see?

- Tom Hanks and Scarlett Johansson
- Scarlett Johansson and Brad Pitt
- Tom Hanks and Brad Pitt
- Scarlett Johansson and Angelina Jolie
- Tom Hanks and Angelina Jolie

B.1.5 Pizzas

The source for this domain is a Montreal pizza restaurant. All these pizzas are either 12 or 13 dollars.

Which one of the following pizzas would you choose?

- Mozzarella, tomato sauce, basil
- Pepperoni, mushrooms, green pepper, mozzarella, tomato sauce
- Red onion, tomato sauce, feta, mozzarella, olive oil, Greek spices, tomato sauce
- Bacon, white onion, mozzarella, parmesan, fresh cream, tomato sauce, ground pepper
- Mushrooms, green pepper, mozzarella, tomato sauce

5

B.1.6 Juices

Which one of the following fresh juices would you choose?

- Mango
- Orange
- Apple
- Grapefruit
- Pineapple

B.1.7 Colours

Which one of the following colours do you like best?

- Red
- Purple
- Pink
- Blue
- Green

B.1.8 Colour pairs

The source for this domain is the website “The top tens”, page “Two colors that look good side by side.” The color combinations here are ranked 1, 4, 5, 13 and 14. We chose a selection of five high ranking combinations among which there were many colors in common. Using a similarity measure equal to the number of colours in common between two pairs, there are two doubleton choice sets where the two colour pairs have no colours in common ($\{a, e\}$ and $\{b, d\}$) and eight where the two colour pairs have one colour in common. This gives six tripleton pairs in which one might expect a similarity effect.

Which one of these colour combinations do you like best?

- Black and red
- Black and purple
- Black and blue
- Blue and red
- Blue and purple

B.1.9 Events

This domain involves comparisons of the probabilities of future events. Logically, the probability of event e must be as least as great as the probability of a, which must in turn be as least as great as the probability of d; also, the probability of b must be at least as great as the probability of c. There is potential for several asymmetric dominance effects in this domain, based on these dominance relations. These would be unconventional effects, as most experimental designs in the

literature intended to elicit asymmetric dominance effects feature numerical dominance relations.

Which one of the following events do you think is most likely to happen in the next twenty years?

- Scotland becomes an independent country.
- Either Catalonia or Quebec become independent countries.
- Catalonia becomes an independent country.
- Scotland and Quebec become independent countries.
- Either Scotland or Quebec become independent countries.

B.1.10 Radio formats

The domain is radio formats, and the choice objects are the top 5 radio formats in Canada in 2015,

5 according to

<https://byrnesmedia.com/2015/10/05/the-6-best-performing-radio-formats-in-canada/>

Descriptions of formats are from

<http://www.newsgeneration.com/broadcast-resources/guide-to-radio-station-formats/>

Suppose you were on a two hour road trip and you have a choice among radio stations with the following formats. Which one would you choose?

- News
- Hot Adult Contemporary, or Hot AC (A variety of classic and contemporary mainstream music geared towards adults.)
- Classic Hits (Rock and pop, roughly 1964-1989)
- Country Music
- Adult Contemporary, or AC (Adult-oriented pop/rock with no hard rock.)

10 B.1.11 Musical artists

The choice objects in this domain are the top selling musical artists of all time, according to Wikipedia. They should be familiar to a large majority of respondents.

Which one of the following musical artists do you like the best?

- The Beatles
- Elvis Presley
- Michael Jackson
- Madonna
- Elton John

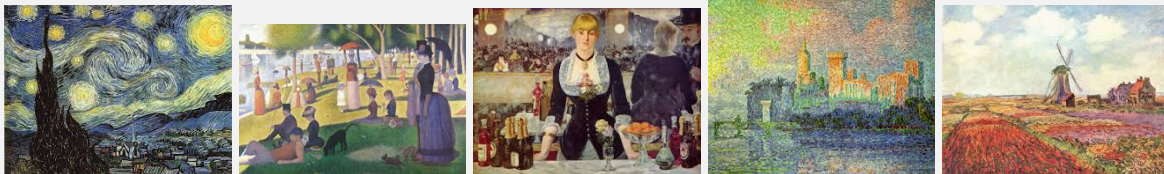
B.1.12 Aboriginal art

Which one of the following examples of Australian aboriginal art do you like the best?



B.1.13 Impressionist art

Which one of the following examples of Impressionist art do you like the best?



B.1.14 Sentences

This domain consists of sentences that have been used in experiments of linguistic judgement. They are, respectively sentences 7j, 7h, 7p, 7a and 7e in Bard et al. (1996). Figure 1 of that paper shows acceptability scores for these and other sentences given by two individual linguists, an acceptability score aggregating the scores of four linguists and an acceptability score aggregating the scores of four “naive respondents”, all undergraduate anatomy students. There is broad, but not perfect, agreement in terms of order, and in the following list, they are in decreasing order of acceptability according to the measure aggregating the judgements of four linguists.

Which one of the following sentences do you find the most grammatically acceptable?

- Who did Bill buy the car to please?
- This is a book which reading would be fun.
- Where did Bill buy the car to drive?
- Which man do you wonder when to meet?
- With which pen do you wonder what to write?

B.1.15 Travel

The source is Tripadvisor. These are the top five travel destinations, according to the results of an on-line contest where visitors to a Tripadvisor site could make pairwise choices between travel

5 destinations.

Which one of the following travel destinations would you most like to visit?

1. Marrakech, Morocco



2. Istanbul, Turkey



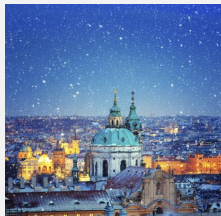
3. Hanoi, Vietnam



4. Siem Reap, Cambodia



5. Prague, Czech Republic



B.1.16 Marijuana

This question elicits policy preferences.

Which one of the following marijuana policies would you choose?

- Possession by, and sales to adults are both legal; sales to minors are illegal.
- Possession by, and sales to adults are both illegal but neither is a criminal offense; sales to minors are a criminal offense.
- Possession is illegal but not criminal; all sales are a criminal offense.
- Possession and sales are criminal offenses, with a small number of medical exceptions.
- Possession and sales are criminal offenses, without exception.

B.2 Single attributes, not directly observed

The five domains in this section contain choice objects with an objective rank order.

B.2.1 Latitude

5 The five cities of this domain have a latitude close to 50 degrees north. In the following list, they are ordered from furthest north to furthest south. According to Wikipedia, their latitudes are, respectively, $52^{\circ}14'N$, $51^{\circ}30'N$, $49^{\circ}15'N$, $48^{\circ}51'N$ and $47^{\circ}36'N$. There are two potential asymmetric dominance effects, with Vancouver being fairly obviously north of Seattle and London being fairly obviously north of Paris.

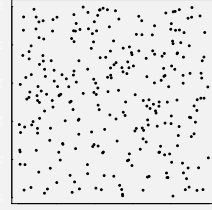
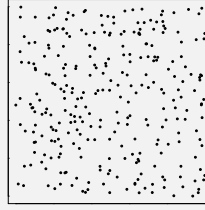
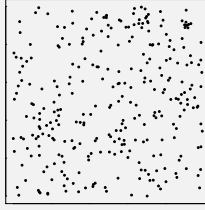
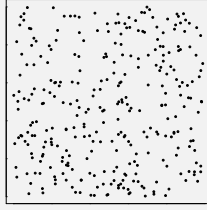
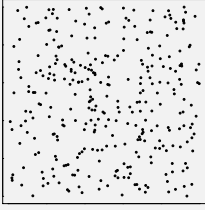
Which one of the following cities do you think is furthest north?

- Warsaw, Poland
- London, United Kingdom
- Vancouver, Canada
- Paris, France
- Seattle, United States

B.2.2 Dots

This domain is a perception example. The true numbers of points are, respectively, 320, 310, 300, 290 and 280. It is much clearer that there are more points in the first panel than in the fifth, than that there are more points in the first than in the second. The difference in the number of points
15 is an obvious similarity measure here that might be expected to lead to similarity effects.

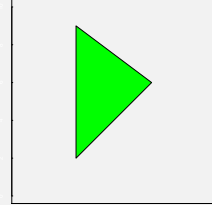
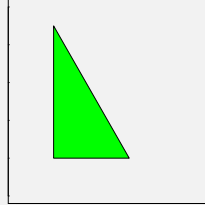
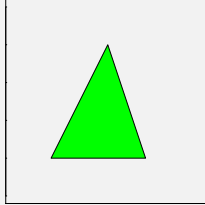
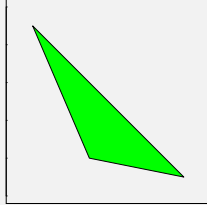
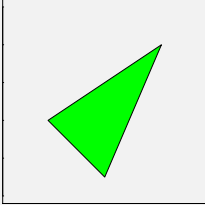
Which one of the following boxes do you think has the greatest number of points?



B.2.3 Triangles

This domain is another perception example. The true areas are, respectively, 16, 15, 15, 14 and 14 units.

Which one of the following triangles do you think has the greatest area?



5

B.2.4 Population

These countries were ranked, respectively, 4th through 8th in terms of population in 2016, when their populations, in millions, were 258, 206, 202, 186 and 156.

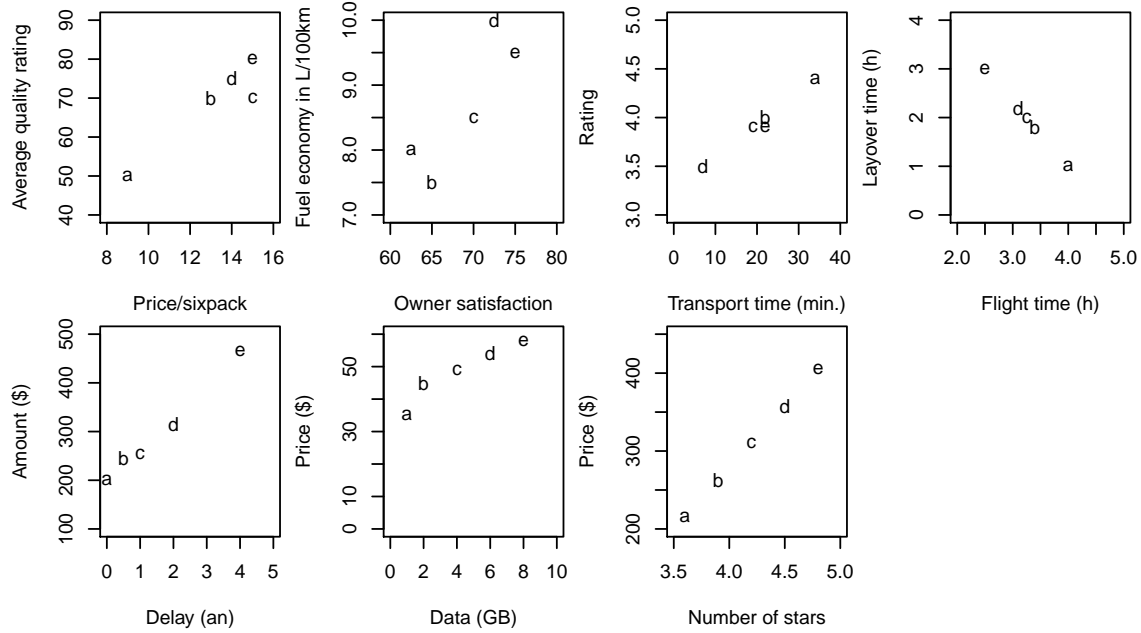
Which one of the following countries do you think has the largest population?

- Indonesia
- Brazil
- Pakistan
- Nigeria
- Bangladesh

10 B.2.5 Surface area

These countries are ranked, respectively, 2nd through 6th in terms of surface area, including inland bodies of water. In millions of square kilometres, those surface areas are 9.984, 9.573, 9.525, 8.516

Figure 6: Designs for context effects. Each panel plots the five objects of a domain in attribute space. The domains are, in row major order, Beer, Cars, Restaurants, Flights I, Delayed Choice, Phone Plans, Hotels



and 7.692.

Which one of the following countries do you think has the greatest surface area, including inland bodies of water?

- Canada
- United States of America
- China
- Brazil
- Australia

B.3 Objects with two attributes

Here we describe domains whose choice objects have exactly two real-valued attributes. In all cases, the levels of these attributes are presented directly to the participant. Figure 6 plots the five objects of each of these domains in attribute space, revealing potential conditions for context effects.

Table 7 shows, for the first four of these domains, dominance, similarity and betweenness relations that we might expect to generate context effects. Following usual practice, we exclude similarity relations for two objects in which there is also a dominance relation between the two;

Domain	Dominance	Similarity w/o dominance	Betweenness w/o similarity
Beer	$e > c, b > c, d > c$	$b \sim d, d \sim e, b \sim e$	$a b e$
Cars	$a > d, b > e$		$a c b, a c e, a c b, d c e$
Restaurants	$b > e, c > e$	$b \sim c$	$a b d, a c d, a e d$
Flights I		$b \sim c, c \sim d, b \sim d$	$a b e, a c e, a d e$

Table 7: Relations of dominance, similarity and betweenness in two-attribute domains

and betweenness relations among three objects when two of them are similar and the third is not similar to either of the first two.

B.3.1 Beer

This domain is from an experiment reported in Huber et al. (1982) used to illustrate an asymmetric dominance effect. The prices are multiplied by 5 and we added choice objects d and e to allow for two more asymmetric dominance effects and two similarity effects. The first panel of Figure 6 show the choice objects in attribute space. Table 7 shows the relations of dominance, similarity and betweenness among objects associated with context effects. The most commonly used domains to illustrate the attraction effect involve Beer, Cars, Apartments, Computers, Restaurants and Televisions.

Below you will find three brands of beer. You know only the price per sixpack and the average quality ratings made by respondents in a blind taste test. Given that you had to choose one brand to buy on this information alone, which one would you choose?

Price/sixpack	Average quality rating (100 = Best; 0 = Worst)
\$9.00	50
\$13.00	70
\$15.00	70
\$14.00	75
\$15.00	80

B.3.2 Cars

This domain is based on an experiment from Wedell & Pettibone (1996). There are two experiments involving cars in that article, the experiment in question is numbered 18 in the appendix to that paper. Objects below have the same attributes as in that experiment and a similar range of levels. We adapted the objects to allow for compromise effects. The second panel of Figure 6 shows

the choice objects in attribute space. Table 6 shows the relations of dominance, similarity and betweenness among objects associated with context effects.

Which one of the following cars would you choose to drive, all other features begin equal?

Ride quality is a on a scale of 0 to 100.

Ride quality	Miles per gallon
60	30
80	24
70	27
55	28
75	22

B.3.3 Restaurants

5 This domain is based on another experiment in Wedell & Pettibone (1996), numbered 19 in the appendix to that paper. The third panel of Figure 6 shows the choice objects in attribute space. Table 7 shows the relations of dominance, similarity and betweenness among objects associated with context effects.

Which one of the following restaurants would you choose for your next restaurant meal, based on transportation time (in minutes) and average customer ratings (from 1 to 5).

Transportation time	Rating
34	4.4
22	4.0
19	3.9
7	3.5
22	3.9

10 **B.3.4 Layovers**

The fourth panel of Figure 6 shows the choice objects in attribute space. Table 7 shows the relations of dominance, similarity and betweenness among objects associated with context effects.

Which one of the following flight itineraries would you choose? All involve two flights, with one layover between them.

Total inflight time	Layover time	Total itinerary time
4:00	1:00	5:00
3:24	1:48	5:12
3:15	2:00	5:15
3:06	2:12	5:18
2:30	3:00	5:30

B.3.5 Delayed Choice

This domain is loosely based on an experiment by Benzion et al. (1989), in which respondents are asked to assign present values equivalent to the receipt of \$200 at time horizons of 0.5, 1, 2 and 4 years. Based on implied discount factors at various terms, we constructed five choice objects designed to have approximately the same present value equivalent.

Which one of the following would you choose?

- \$200 credited to your bank account immediately.
- \$245 credited to your bank account in six months.
- \$255 credited to your bank account in one year.
- \$315 credited to your bank account in two years.
- \$465 credited to your bank account in four years.

B.3.6 Phone plans

The source for this domain is the website of Fido Mobile, a Canadian cell phone carrier, with rates quoted on March 1, 2017 in Canadian dollars.

Of the following cell phone plans, which one would you choose? In all cases, unlimited calling, text picture and video messages to Canadian and international mobile numbers are included. Excess data usage is billed at \$10 per 500 MB.

- 1 GB data per month, \$35 per month.
- 2 GB data per month, \$45 per month.
- 4 GB data per month, \$49 per month.
- 6 GB data per month, \$54 per month.
- 8 GB data per month, \$58 per month.

B.3.7 Hotel rooms

Using Expedia results, I did a linear regression of price per night on a constant and the Expedia rating, in numbers of stars, for a sample of available hotels. The five levels of numbers of stars
5 correspond roughly to the mean, plus and minus one sample standard deviation and plus and minus two standard deviations. Prices are approximately equal to fitted values in the linear regression.

Suppose you are staying over two nights in New York city. Which one of the following hotels would you choose, based on customer ratings and price per night?

- 3.6/5 stars, \$215 per night
- 3.9/5 stars, \$263 per night
- 4.2/5 stars, \$311 per night
- 4.5/5 stars, \$358 per night
- 4.8/5 stars, \$406 per night

B.4 Objects with multiple attributes

B.4.1 Flight itineraries

- 10 This domain illustrates three-way tradeoffs. The points form a constellation in the simplex that resembles the pattern of points on the “five” side of a die.

Which one of the following flight itineraries would you choose? All involve two flights and have a total duration of six hours.

1st flight	Layover	2nd flight
1:30	1:15	3:15
3:15	1:15	1:30
2:15	1:30	2:15
1:30	1:45	2:45
2:45	1:45	1:30

B.4.2 Televisions

The source for this domain is the website of Best Buy Canada, with prices in Canadian dollars.

Which one of the following televisions would you choose to buy if you were in the market for a television? All are LED televisions. Resolution refers to number of horizontal lines. Smart indicates internet connectivity.

Brand	Resolution	Smart	Price (\$)	Screen Size (inches)
Sharp	1080	Yes	309	32
Insignia	720	No	209	32
Sony	720	Yes	439	32
Samsung	1080	Yes	459	40
Toshiba	1080	No	409	43

5 B.4.3 Coffee

The source for this domain is the website buycoffeecanada.com, with prices in Canadian dollars.

You need to buy 16oz of ground coffee for a brunch with friends. Which one of the following ground coffees would you choose?

Price	Fair Trade	Name: Description
18.71	Yes	Ethiopian Yirgacheffe: vibrant and intensely aromatic, fruity
9.99	No	Colombian Supremo: mellow cup, complex aromas and rich flavours
13.72	Yes	Colombian Organic: medium body, fragrant aroma and mild acidity
12.35	No	Tanzania Peaberry: full bodied coffee, chocolatey aroma, wine-like finish
13.46	No	Sumatra Mandheling: exotic, earthy, bright with low acidity

B.4.4 Charity

The charities in this domain are real charities. The first two are relatively innocuous in the sense that most people support the goals of both. However, the Canadian Cancer Society attracts much more financial support than Arthritis Research Canada and we might expect that most people would prefer a marginal dollar going to the former. The two other charities have goals that are nearly opposite and elicit strong emotions (of different kinds) from some.

Suppose someone was donating a total of 100 dollars to a combination of charities, on your behalf. Which one of the following divisions of the 100 dollars would you choose?

Arthritis Research Canada	Canadian Cancer Society	Canadian Coalition for Firearm Rights	Coalition for Gun Control
90	10	0	0
35	60	5	0
35	60	0	5
10	80	10	0
10	80	0	10

Statements and Declarations

Authors' Contribution Statement

McCausland contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, software, visualization and writing (initial draft, review and

editing). Marley contributed to the conceptualization, funding acquisition, investigation, project administration, and writing-review and editing. Davis-Stober contributed to the conceptualization, investigation, and writing - review and editing.

Acknowledgements

- 5 The authors thank the Institute for Choice at the University of South Australia for programming and hosting the survey, Survey Sampling International for recruiting participants, and the participants themselves. They also thank audiences at various conferences for insightful questions and helpful suggestions.

Competing Interests

- 10 The authors declare none.

Financial Support

This work was supported by the Social Sciences and Humanities Research Council of Canada under Insight grant SSHRC 435-2012-0451 to the University of Victoria for Marley and McCausland.

References

- 15 Bard, E. G., Robertson, D. & Sorace, A. (1996), ‘Magnitude estimation of linguistic acceptability’, *Language* **72**(1), 32–68.
- Benzion, U., Rapoport, A. & Yagil, J. (1989), ‘Discount rates inferred from decisions: An experimental study’, *Management Science* **35**, 270–84.
- Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis, Second Edition*, Springer-
20 Verlag, New York, NY.
- Bernardo, J. M. & Smith, A. F. M. (1994), *Bayesian Theory*, John Wiley and Sons, Chichester, England.
- Block, H. D. & Marschak, J. (1960), Random orderings and stochastic theories of responses, *in*
I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow & H. B. Mann, eds, ‘Contributions to
25 Probability and Statistics: Essays in Honor of Harold Hotelling’, Stanford University Press, Stanford, CA, pp. 97–132.

- Cavagnaro, D. R. & Davis-Stober, C. P. (2014), ‘Transitive in our preferences, but transitive in different ways: An analysis of choice variability’, *Decision* **1**, 102–122.
- Davis-Stober, C. P., Brown, N. & Cavagnaro, D. R. (2015), ‘Individual differences in the algebraic structure of preference’, *Journal of Mathematical Psychology* **66**, 70–82.
- 5 de Condorcet, L. M. (1785), *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix.*, L’imprimerie Royale, Paris.
- Falmagne, J. C. (1978), ‘A representation theorem for finite random scale systems’, *Journal of Mathematical Psychology* **18**, 52–72.
- Flegal, J. M. & Jones, G. L. (2010), ‘Batch means and spectral variance estimators in Markov chain
10 Monte Carlo’, *Annals of Statistics* **38**(2), 1034–1070.
- Huber, J., Payne, J. W. & Puto, C. (1982), ‘Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis’, *Journal of Consumer Research* **9**, 90–98.
- Kass, R. E. & Raftery, A. E. (1995), ‘Bayes factors’, *Journal of the American Statistical Association* **90**, 773–795.
- 15 Kelley, T. L. (1928), *Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities*, Stanford University Press.
- Klugkist, I. & Hoijsink, H. (2007), ‘The Bayes factor for inequality and about equality constrained models.’, *Computational Statistics and Data Analysis* pp. 6367–6379.
- Krantz, D. H. (1967), ‘Rational distance function for multidimensional scaling’, *Journal of Mathematical Psychology* **4**, 226–245.
20
- Luce, R. D. & Suppes, P. (1965), Preference, utility, and subjective probability, in R. D. Luce, R. R. Bush & E. Galanter, eds, ‘Handbook of Mathematical Psychology’, Vol. 3, John Wiley & Sons, New York, NY, chapter 19, pp. 249–410.
- McCausland, W. J., Davis-Stober, C. P., Marley, A. A. J., Park, S. & Brown, N. (2020), ‘Testing
25 the random utility hypothesis directly’, *The Economic Journal* **130**, 183–207.
- McCausland, W. J. & Marley, A. A. J. (2013), ‘Prior distributions for random choice structures’, *Journal of Mathematical Psychology* **57**, 78–93.

- McCausland, W. J. & Marley, A. A. J. (2014), ‘Bayesian inference and model comparison for random choice structures’, *Journal of Mathematical Psychology* **62-63**, 33–46.
- McCausland, W. J. & Marley, A. A. J. (2024), ‘On the multiplicative inequality’, *Journal of Mathematical Psychology* . forthcoming.
- 5 Myung, J. I., Karabatsos, G. & Iverson, G. J. (2005), ‘A Bayesian approach to testing decision making axioms’, *Journal of Mathematical Psychology* **49**, 205–225.
- Sattath, S. & Tversky, A. (1976), ‘Unite and conquer: A multiplicative inequality for choice probabilities’, *Econometrica* **44**, 79–89.
- Simonson, I. (1989), ‘Choice based on reasons: The case of attraction and compromise effects’,
 10 *Journal of Consumer Research* **16**.
- Train, K. (1978), ‘A validation test of a disaggregate mode choice model’, *Transportation Research* **12**, 167–174.
- Tversky, A. (1972a), ‘Choice by elimination’, *Journal of Mathematical Psychology* **9**, 341–367.
- Tversky, A. (1972b), ‘Elimination by aspects: A theory of choice’, *Psychological Review* **79**(4), 281–
 15 299.
- Tversky, A. & Simonson, I. (1993), ‘Context-dependent preferences’, *Management Science* **39**, 1179–1189.
- Wedell, D. H. & Pettibone, J. C. (1996), ‘Using judgments to understand decoy effects in choice’, *Organization Behavior and Human Decision Processes* **67**(3), 326—344.
- 20 Zwilling, C., Cavagnaro, D. & Regenwetter, M. (2011), Quantitative testing of decision theories: A Bayesian counterpart. Presentation at the Annual Meeting of the Society for Mathematical Psychology, Boston, July 15, 2011.