

# A choice experiment for testing axioms of stochastic discrete choice

William J. McCausland<sup>\*1</sup>, A.A.J. Marley<sup>2</sup> and Clinton P. Davis-Stober<sup>†3</sup>

<sup>1</sup>Département de sciences économiques and CIREQ, Université de Montréal, Canada

<sup>2</sup>Department of Psychology, University of Victoria, Canada

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

August 28, 2025

## Abstract

We run an experiment covering 32 different choice domains, including fine art, cars and  
pizza. For each domain, there is a set of 5 related objects, and 26 menus—the subsets of two  
or more objects. We measure population level choice probabilities by presenting each menu  
to 40 participants and observing their choices of an object. The point of the experiment is to  
test axioms of stochastic discrete choice for population level choice probabilities, and it is well  
suited for testing axioms, like regularity, random utility and the multiplicative inequality, which  
constrain choice probabilities across all menus of a domain. Each participant faces exactly one  
menu from each domain; for between-subject designs like this, the assumption that choices are  
independent across trials, and on each menu, identically distributed, is plausible. Participants  
are representative of the Canadian adult population. For each domain and each of the three  
axioms above, we report Bayes factors giving evidence for or against the axiom. Evidence sup-  
ports regularity and random utility, for each domain; evidence for the multiplicative inequality  
is mixed.

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

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

---

<sup>\*</sup>Corresponding author, [william.j.mccausland@umontreal.ca](mailto: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>.

<sup>†</sup>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, as the case may be, 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. 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* indistinguishable: before observing choices among objects, all we know about an object is its index, allowing us to follow the object across all the menus that include it. In different analyses, choice data observed over several trials may be a sequence of choices made a single decision maker or the choices of a sample 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, cars and pizza. In this way, we can see which choice patterns are more universal and which are more idiosyncratic. Second, for each domain, we observe choices from every menu; that is, from each set of at least two objects from the domain. Each domain has five objects, and thus<sup>1</sup> 26 menus. For each menu of each domain, we observe the choices of 40 different participants, allowing us to measure population choice probabilities with some precision. For axioms such as regularity, which constrains choice probabilities across all menus of a domain, this allows us to expose all implications of the axiom to possible falsification. Third, the experiment features a strictly between-subject design, where each participant faces exactly one menu from each domain. For these designs, the standard assumption that choices are independent across trials, and on each menu identically distributed, is plausible.

---

<sup>1</sup>From the  $2^5 = 32$  subsets of a set of 5 objects, we subtract 5 singleton sets and the empty set.

We provide a preliminary empirical analysis of these new data using Bayesian methods developed in two previous papers. McCausland & Marley (2013) develops non-parametric models, defined up the choice of a prior distribution, for systems of choice probabilities. The paper also provides tools for detecting whether or not a given system of probabilities satisfies various choice axioms. 5 McCausland & Marley (2014) provides methods for the simulation of choice probabilities from their posterior distribution, for these non-parametric models. It also describes methods for testing various axioms using Bayes factors. Together, the two papers establish a testing ground for stochastic choice axioms that is extendable to other axioms. Our preliminary analysis consists of tests of three axioms for which the new data are particularly well suited: regularity, random utility and 10 the multiplicative inequality. We do this for each of the 32 choice domains of the experiment.

## 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 system of choice 15 probabilities  $P_A(x)$ ,  $x \in A \subseteq T$ , and is denoted  $P$ . We will write  $T = \{a, b, c, d, e\}$  for a generic universe with five elements; for the “Colours” domain of the experiment,  $a$  is Red,  $b$  is Purple,  $c$  is Pink,  $d$  is Blue, and  $e$  is Green. One of the 26 menus is  $\{a, b\}$ ;  $P_{\{a,b\}}(a)$  and  $P_{\{a,b\}}(b)$  are the probabilities of choosing Red and Purple from the menu consisting of these two options. We learn something about these probabilities from the experiment: 21 participants chose Red and 19 chose 20 Purple, as we can see in the first row of Table 4 of Section 5, which tabulates choice counts for each of the 26 menus of the “Colours” domain. Denote the observed proportion of times that an object  $x$  is chosen from a menu  $A$ ,  $x \in A \subseteq T$ , by  $\hat{P}_A(x)$ . For example, we compute the observed choice proportions  $\hat{P}_{\{a,b\}}(a) = 0.525$  and  $\hat{P}_{\{a,b\}}(b) = 0.525$  for the “Colours” domain.

Let  $\Delta$  be the space of random choice structures;  $\Delta$  is a subset of Euclidean space defined by the 25 non-negativity of choice probabilities and a set of equalities requiring choice probabilities on each menu to add up to one.

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 30 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 assume 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 the iid assumption 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. Historically, many of these axioms pertain only to binary choice probabilities. See, for example, Luce & Suppes (1965) for the various forms of stochastic transitivity and also the triangle inequality.

Given the nature of the experimental data we introduce in this paper, we will be more interested in the following three axioms, each of which relates choice probabilities across all menus of a choice universe, not just the doubleton menus. A random choice structure  $P$  satisfies

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

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

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

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

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

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

If a random choice structure  $P$  for the “Colours” domain satisfies regularity, then the probability of choosing Pink from a menu consisting of Red, Purple and Pink cannot be greater than the probability of choosing Pink from the doubleton menu of Purple and Pink. The fact that six  
 5 participants out of forty chose Pink from the first menu and only three participants out of forty chose Pink from the smaller menu means that the observed system of choice proportions violates regularity. However, before rejecting the hypothesis that the population RCS violates regularity, we have to take into account sampling variability and the fact that the regularity condition consists of multiple inequalities. Luce & Suppes (1965) describe regularity and show that it is necessary for  
 10 random utility.

Block & Marschak (1960) show that the Block-Marschak inequalities are necessary for random utility and Falmagne (1978) shows they are sufficient.

If  $P$  satisfies the multiplicative inequality, then the probability of choosing Purple from the tripleton menu of Red, Purple and Pink has to be at least as great as the joint probability of  
 15 choosing Purple over Red in pairwise choice and then Purple over Pink in pairwise choice. In the experiment, 15 out of 40 participants chose Purple from the tripleton menu, 19 out of 40 chose Purple over Red and 37 out of 40 chose Purple over Pink. It is easy to check that the choice proportions violate the multiplicative inequality. The multiplicative inequality is due to Sattath & Tversky (1976), who show that it is a testable implication of the independent random utility  
 20 model and also of the Elimination by Aspects (EBA) model described in Tversky (1972a) and Tversky (1972b). McCausland & Marley (2024) show that any system of binary probabilities can be extended to a full random choice structure in such a way that the random choice structure satisfies the multiplicative inequality. They also prove several results on the relationship between regularity and the multiplicative inequality. See Luce & Suppes (1965) and McCausland & Marley  
 25 (2013) for illustrations of the logical relationships among a set of axioms that includes these three and various kinds of stochastic transitivity.

We can immediately see from the definitions that for each of these axioms, every choice probability vector  $P_A$  appears in many constraints. We can see moreover that each  $P_A$  is *pivotal*, in the sense that for any given random choice structure satisfying the axiom in question and an arbitrary

$A$ , one can change  $P_A$  to make the random choice structure violate the axiom. It is for this reason that to test an axiom such as these, it is important to observe choice data from every menu.

It will be helpful to frame the three axioms as defining subsets of the space  $\Delta$  of random choice structures. Let  $\Lambda_{\text{Reg}}$  be the subset of  $\Delta$  where regularity holds; similarly, define  $\Lambda_{\text{RU}}$  and  $\Lambda_{\text{MI}}$  as the subsets of  $\Delta$  where random utility and the multiplicative inequality hold. Since regularity and random utility are collections of linear inequalities in probabilities and  $\Delta$  is convex,  $\Lambda_{\text{Reg}}$  and  $\Lambda_{\text{RU}}$  are both convex. McCausland & Marley (2024) show that  $\Lambda_{\text{MI}}$  is not convex, regardless of the number of objects in the choice universe.

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. Nonetheless, for “convex” axioms such as regularity and random utility, if the RCS for all individuals in a population satisfy the axiom, then the population RCS must as well. This is because the population RCS, being a convex combination of individual RCSs, must satisfy the axiom whenever all the individual ones do. Expressed in contrapositive form to make plain the implications for testing, if the axiom is violated for a population, than it must be violated by at least one individual.

However, there are other axioms that can hold for every individual in a population but not the population itself, including the multiplicative inequality. We provide such a counterexamples where a set of individual RCSs satisfying an axiom (or axioms) aggregate to population RCSs that do not. 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 a counterexample. Lines 1 and 2 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 3, which violates the multiplicative inequality since

$$P_{\{x,y,z\}}(y) = \frac{1}{2} < P_{\{y,x\}}(y)P_{\{y,z\}}(y) = \left(\frac{3}{4}\right)^2 = \frac{9}{16}.$$

### 1.3 Context effects

One of the 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 A, where we describe the choice domains in detail, we describe how some of the domains are designed to elicit these effects.

Line	$p_i$	$\pi_{xyz}$	$\pi_{xzy}$	$\pi_{yxz}$	$\pi_{yzx}$	$\pi_{zxy}$	$\pi_{zyx}$	$P_{\{x,y\}}(x)$	$P_{\{y,z\}}(y)$	$P_{\{x,z\}}(x)$	$P_T(x)$	$P_T(y)$
1	$\frac{1}{2}$	$\frac{1}{2}$		$\frac{1}{2}$				$\frac{1}{2}$	1	1	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{2}$			$\frac{1}{2}$			$\frac{1}{2}$		$\frac{1}{2}$			$\frac{1}{2}$
3		$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$

Table 1: An aggregation of individual choice probabilities. Each line gives a random ranking (in columns  $\pi_{xyz}$  through  $\pi_{zyx}$ , where for example  $\pi_{xyz}$  is the probability of the ranking  $x, y, z$ ) and the implied RCS (columns  $P_{\{x,y\}}(x)$  through  $P_T(y)$ ). The individual random rankings and RCS's in lines 1 and 2 aggregate (with mixing probabilities  $p_i$ ) to the population random ranking and RCS in line 3.

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. It 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  $P_{\{x,y,z\}}(x) > P_{\{x,y\}}(x)$ , 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

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

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:

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

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

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 distribution  $\mathbf{X}|P$ .

A *Bayesian discrete choice model* consists of a prior distribution over  $\Delta$ . 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  $\Delta$ , the mixture distribution is the marginal probability mass function  $f_M(\mathbf{x}) = \int_{\Delta} f(\mathbf{x}|P) f_M(P) dP$ . Evaluation of  $f_M(\mathbf{x})$  at the observed value  $\mathbf{x}_{\text{obs}}$  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  $\mathbf{x}_{\text{obs}}$ . 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  $\mathbf{x}_{\text{obs}}$ .

To test axioms, we will compare *encompassing* models, Bayesian discrete choice models where the prior density  $f(P)$  has full support on  $\Delta$ , 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  $\Lambda_{\text{SST}}$ , for example, is the subset of  $\Delta$  where strong stochastic transitivity holds. A constrained model is a Bayesian discrete model whose prior density over  $\Delta$  is the truncation of the prior density in an encompassing model to  $\Lambda$ . 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. (2015), Myung et al. (2005) and Zwillling 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 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 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  $\alpha$  describes how consistent choices are likely to be across trials. For each menu  $A = \{a_1, \dots, a_{|A|}\}$ , the marginal distribution of the vector  $(P_A(a_1), P_A(a_2), \dots, P_A(a_{|A|}))$  is symmetric Dirichlet with all  $|A|$  shape parameters set to  $\alpha/|A|$ . Figure 1 shows densities of binary choice probabilities, for six different values of  $\alpha$ . For  $\alpha < |A|$ , the density of  $P$  is unbounded. We see in Figure 1, for example, that for  $\alpha = 1$ , the density of any binary choice probability approaches infinity as the probability goes to zero or one. As  $\alpha$  gets close to zero, more and more probability is concentrated in regions where probabilities are close to zero or one, and choice looks more and more deterministic. For  $\alpha > |A|$ ,  $(P_A(a_1), P_A(a_2), \dots, P_A(a_{|A|}))$  has a mode at  $(1/|A|, \dots, 1/|A|)$  and the density approaches zero as any probability goes zero. As  $\alpha$  gets very large, more and more probability is concentrated at this uniform-discrete-choice mode and choice looks more and more unpredictable. Figure 1 shows that as  $\alpha$  increases from 5 to 10 to 20, the mode (and mean) of any binary choice probability stays at  $\frac{1}{2}$  but its variance decreases. For the intermediate case  $\alpha = |A|$ ,  $P_A(\cdot)$  is uniformly distributed on the probability simplex.

The parameter  $\lambda$  describes the degree of dependence of choice probabilities across menus. For

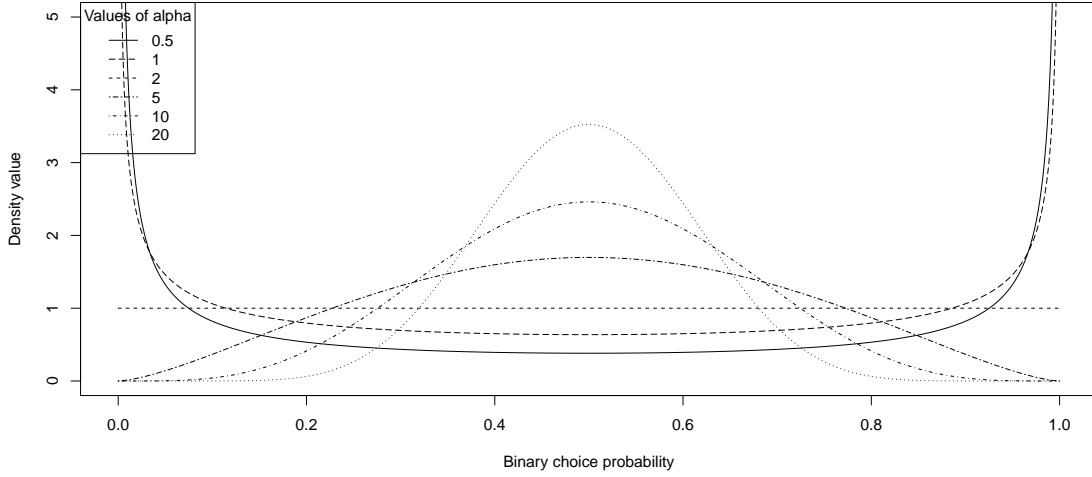


Figure 1: Density of binary choice probabilities for  $\alpha = 0.5, 1, 2, 5, 10, 20$

$\lambda = 0$ , 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  $\Delta$ . Details on the prior distributions for  $P$ , including some of its desirable properties, are given in McCausland & Marley (2013).

It remains to specify the unknown parameters  $\alpha$  and  $\lambda$ . One can choose fixed values, but a more flexible approach allows for learning about plausible values of  $\alpha$  and  $\lambda$  through data. In this approach, one provides a prior distribution for  $\alpha$  and  $\lambda$ , 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  $\alpha$ ,  $\lambda$  and  $P$  using a hierarchical prior consisting of a prior density  $f_M(\alpha, \lambda)$  for  $\alpha$  and  $\lambda$  and the density  $f(P|\alpha, \lambda)$  in McCausland & Marley (2013). 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  $\mathbf{x}_{\text{obs}}$ .

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  $(\alpha, \lambda)$  where  $\alpha$  and  $\lambda$  are *a priori* independent,  $\alpha$  has a Gamma distri-

	$a_1$	$a_2$	$b$	$E[\alpha]$	$\text{Var}[\alpha]$	$\sigma_\alpha$	$E[\lambda]$	$\text{Var}[\lambda]$	$\sigma_\lambda$
$M_{\text{ind}}$	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

bution with shape parameter  $a_1 + a_2$  and rate parameter  $b$ ,  $\lambda$  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  
5 paper is concerned with population, rather than individual choice. Anticipating some heterogeneity across individuals, we shift prior weight towards higher values of  $\alpha$ , 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  $\alpha$ , and that values of  $\lambda$  close to one were strongly favoured. Accordingly, we use the same prior for  $\alpha$  for each encompassing model and put more  
10 prior weight on values of  $\lambda$  closer to one than in McCausland & Marley (2014) and McCausland et al. (2020).

Table 2 shows the values of  $(a_1, a_2, b)$  for four encompassing models used in the empirical analysis. Model  $M_{\text{ind}}$  features a degenerate distribution for  $\lambda$ , 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  
15 less restrictive axioms (WST, MST, TI) as the prior probability of the other axioms under  $M_{\text{ind}}$  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  $\alpha$  and  $\lambda$  for the four encompassing models. As we go from model  $M_1$  to  $M_2$  and then  $M_3$ , the prior distribution of  $\lambda$   
20 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  
25 of  $P$  and the parameters  $\alpha$  and  $\lambda$  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  $\alpha$ ,  $\lambda$  and  $P$ , from the choice dataset  $\mathbf{x}_{\text{opt}}$ . The posterior distribution is analytically intractable and we use Markov chain Monte Carlo (MCMC) methods to obtain a large posterior 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  $\mathbf{x}_{\text{opt}}$  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  $\Lambda$  defining an axiom, then the Bayes factor becomes

$$\frac{f_{M_c}(\mathbf{x}_{\text{opt}})}{f_{M_e}(\mathbf{x}_{\text{opt}})} = \frac{\Pr_{M_e}[\mathbf{X} = \mathbf{x}_{\text{opt}} | P \in \Lambda]}{\Pr_{M_e}[\mathbf{X} = \mathbf{x}_{\text{opt}}]} = \frac{\Pr_{M_e}[P \in \Lambda | \mathbf{X} = \mathbf{x}_{\text{opt}}]}{\Pr_{M_e}[P \in \Lambda]}. \quad (2)$$

The second equation of (2) follows from Bayes' rule; see Klugkist & Hoijsink. (2007) for details; the far right hand side of (2) 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. 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, 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. The Supplementary Materials include copies of the consent form, a Certificate of (ethics) Approval, the SSI Global Panel Book 2017, and a document describing in more detail the recruitment and compensation procedures followed by SSI.

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 or territory, 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. No participant was in Nunavut. 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. A complete list of participants' feedback is in the Supplementary Materials. 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 A describes each domain in more detail, including the question, the five choice objects and, where applicable, a source for the information used to construct the question.

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. Figure 2 shows screenshots for the “Travel” and “Coffee” domains. The screenshot for the “Travel” domain shows a menu with two of the five choice objects in that domain. The screenshot for the “Coffee” domain shows a menu with four of the five choice objects in that domain.

Each participant faces exactly one menu from each domain, so the number of choice trials for each participant is the same as the number of domains,  $J = 32$ . Regarding the assignment of participants to menus, our intended design was as follows. For each domain, each of the  $I = 26$  menus of that domain is presented  $N = 40$  times, once to each of  $N$  distinct participants. So for the “Colours” domain, for example,  $N = 40$  different participants see the menu {Green, Blue} and another  $N$  see the menu {Blue, Red, Purple}, and so on for all the menus of that domain. With  $I = 26$  menus in total, we have  $N I = 40 \cdot 26 = 1040$  distinct participants, each choosing one colour from some menu of colours. For another domain, the same 1040 participants all choose one object from some menu of that other domain and again, each of the  $I = 26$  menus of the new domain is seen by  $N = 40$  different participants.

Thus the total number of trials is  $N I J = 40 \cdot 26 \cdot 32 = 33,280$ , which we can factor in different ways. Each of the  $N I = 1040$  participants sees one menu from each of the  $J = 32$  domains for a total of  $N I \cdot J$  trials. Each of the  $J = 32$  domains has  $I = 26$  menus and each of these  $J I = 32 \cdot 26 = 832$  menus is seen by  $N = 40$  participants, for a total of  $J \cdot I \cdot N$  trials.

Assignments of participants to menus() and the positions of menu items on the screen are fully randomized. 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, so it is unlikely that any two


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 combos	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?
Aboriginal	Which one of the following examples of Australian aboriginal art do you like the best?
Impressionist	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?
Surface 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 being equal? ...
Restaurants	Which one of the following restaurants would you choose for your next restaurant meal...?
Layovers	Which one of the following flight itineraries would you choose? ...
Future	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	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?

participants will be in the same group and thus see the same menu() more than a couple of times. Elements of a menu are presented in random position (left to right or top to bottom, as the case may be) on the screen, independently across the  $N$  participants facing that same menu. The  $J$  trials

Travel

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


Istanbul, Turkey


Marrakech, 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.

assigned to each participant are presented in random order, independently across participants, so it is exceedingly unlikely that two participants see the domains in the same order. 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, exactly two menus were presented 41 times and the remaining 24 menus were presented 40 times, as planned. Thus the actual total number of trials is 33,344.

Participants are constrained to 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 experiment.

Data and code are available in the Supplementary Materials.



## 5 Results

We begin by discussing constraints on generality. We then provide 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  
5 “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  $\alpha$  and  $\lambda$  for the various choice domains.

### 5.1 Constraints on generality

We would expect choice proportions for the domains studied here to vary across populations and through time. We suspect, for example, that other populations may be more or less knowledgeable about the latitude of North American and European cities and that tastes in actors evolve over time. We conjecture that the patterns of choice probabilities that we find here, in terms of their adherence  
15 to various axioms of discrete choice, are more universal, but this is ultimately an empirical question.

### 5.2 Engagement of participants

Evidence of 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  
20 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  
25 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.

### 5.3 Observed choice counts and proportions

We illustrate the data for the “Colours” domain. Table 4 reports observed choice counts and corresponding proportions for each menu. Figure 3 then provides a structured graphical representation of the observed proportions for binary and ternary menus. Similar summaries for all 32 domains are available in the Supplementary Materials.

For the “Colours” domain, 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 the response ‘Red’,  $b$  is Purple,  $c$  is Pink,  $d$  is Blue and  $e$  is Green.

Table 4 shows the observed choice counts and choice proportions for all menus. Each row gives counts and proportions for the menu displayed in the first column. The second through sixth columns show the number of participants who chose the objects  $a, b, c, d$  and  $e$ , respectively, from the menu. Where an object is not in the menu, a dash is displayed. The final five columns show the corresponding choice proportions.

Figure 3 is a graphical representation that displays, among other things, choice proportions for all doubleton and tripleton menus of the “Colours” domain. The figure consists of ten panels, one for each tripleton subset of  $\{a, b, c, d, e\}$ . The panel for a given tripleton shows an equilateral triangle with each vertex labelled with one of its elements. Consider the first panel, for the tripleton  $T \equiv \{a, b, c\}$ , which we will use as an extended example. The vertices of the triangle are labelled  $a$  (bottom left),  $b$  (top) and  $c$  (bottom right). There are four menus in this set,  $\{a, b\}$ ,  $\{a, c\}$ ,  $\{b, c\}$  and  $\{a, b, c\}$  and the choice proportions observed for these menus appear in rows 1 through 4 of Table 4, respectively.

We can use an equilateral triangle as a Barycentric coordinate system for probabilities or proportions on a tripleton. For example, an arbitrary probability vector  $(P_A(a), P_A(b), P_A(c))$  on  $T$  is represented as the point on the (solid) triangle with Barycentric coordinates  $(P_A(a), P_A(b), P_A(c))$ . We have represented the observed vector  $(\hat{P}_T(a), \hat{P}_T(b), \hat{P}_T(c)) = (\frac{19}{40}, \frac{15}{40}, \frac{6}{40}) = (0.475, 0.375, 0.15)$  of choice proportions observed in the experiment as a solid point in the interior of the triangle. There are two ways of locating this point on the triangle. First, we can take vertices  $a, b$  and  $c$  as points on the plane of the page or screen. Then the point  $(\hat{P}_T(a), \hat{P}_T(b), \hat{P}_T(c)) = (0.475, 0.375, 0.15)$  is 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 0.475, 0.375 and 0.15. Thus, the solid point is at  $0.475a + 0.375b + 0.15c$ . The second way relies on the fact that the triangle is equilateral. Taking the triangle to have unit height, the point representing  $(\hat{P}_T(a), \hat{P}_T(b), \hat{P}_T(c)) = (0.475, 0.375, 0.15)$  is at a distance of  $\hat{P}_T(a) = 0.475$ , to the right side of the triangle, at a distance of  $\hat{P}_T(b) = 0.375$  to the base and at a distance of  $\hat{P}_T(c) = 0.15$  to the left side. The downward sloping grid lines, the horizontal grid lines, and the upward sloping grid lines measure out the choice probabilities (or proportions) of  $a$ ,  $b$  and  $c$ , respectively.

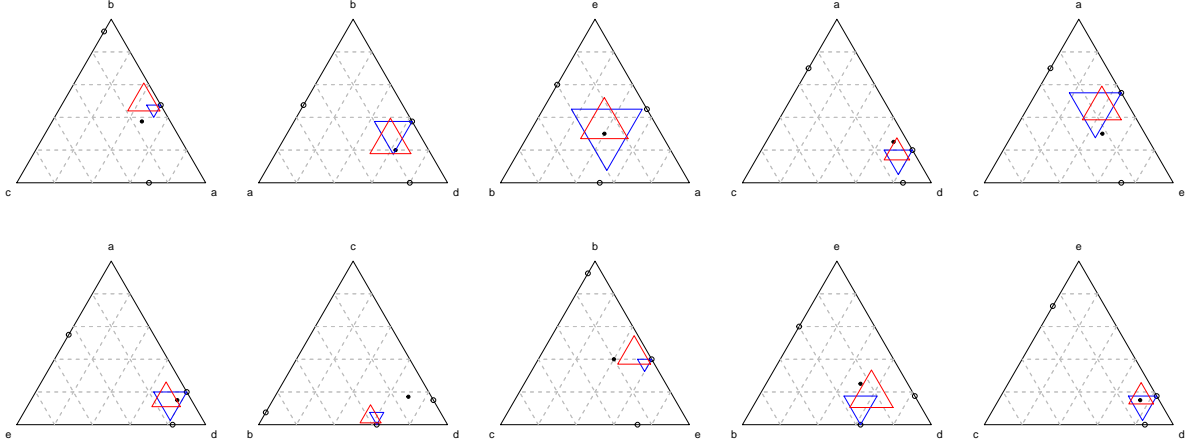
The hollow dots on the left, right and bottom sides of the triangle represent the choice proportions  $\hat{P}_{\{a,b\}}(a) = \frac{21}{40} = 0.525$ ,  $\hat{P}_{\{b,c\}}(b) = \frac{37}{40} = 0.925$  and  $\hat{P}_{\{(a,c)\}}(c) = \frac{28}{40} = 0.7$ , respectively. For example, the hollow dot on the left side of the triangle is the convex combination of the vertices labelled  $a$  and  $b$ , with weights 0.525 and 0.475 respectively. Equivalently, the distance from this hollow dot to vertex  $a$  and the distance to vertex  $b$  are in the proportions 0.525 and 0.475.

The two triangles in the interior allow readers to detect, at a glance, if binary and ternary choice proportions for the tripleton set  $A$  violate regularity or the multiplicative inequality. Suppose that we know the binary choice proportions  $\hat{P}_{\{a,b\}}(a) = \frac{21}{40} = 0.525$ ,  $\hat{P}_{\{b,c\}}(b) = \frac{37}{40} = 0.925$  and  $\hat{P}_{\{(a,c)\}}(c) = \frac{28}{40} = 0.7$ , but we have not yet observed the ternary choice proportions  $(\hat{P}_A(a), \hat{P}_A(b), \hat{P}_A(c))$ .

The downward pointing solid triangle with the blue boundary is a cross-section of the region where regularity holds. For the three given binary choice proportions (indicated by the three hollow dots) the solid triangle is the set of choice proportions consistent with regularity and those binary choice probabilities. It is defined by the regularity conditions on  $T$ , which simplify to  $\hat{P}_T(a) \leq \min(\hat{P}_{\{a,b\}}(a), \hat{P}_{\{a,c\}}(a))$ ,  $\hat{P}_T(b) \leq \min(\hat{P}_{\{b,a\}}(b), \hat{P}_{\{b,c\}}(b))$  and  $\hat{P}_T(c) \leq \min(\hat{P}_{\{c,a\}}(c), \hat{P}_{\{c,b\}}(c))$ . In the event,  $(\hat{P}_T(a), \hat{P}_T(b), \hat{P}_T(c))$  lies outside the triangle with the blue boundary. Specifically,  $\hat{P}_A(c) = 0.15$  is greater than  $\hat{P}_{\{c,b\}} = 0.075$ . We conclude that the binary and ternary choice proportions on  $T$  do not satisfy regularity. It is important to remember that the observed choice proportions are subject to sampling variation and so the fact that choice proportions violate regularity does not constitute a rejection of the hypothesis that population probabilities violate regularity. We are describing here a data visualization device and not the results of statistical tests.

Similarly, the upward pointing solid triangle with the red boundary is a cross-section of the region where the multiplicative inequality holds. For the given binary choice proportions, the solid triangle is the set of ternary choice proportions consistent with the multiplicative inequality and the observed binary choice proportions. It is defined by the multiplicative inequalities on  $T$ , which

Figure 3: Observed binary and ternary choice proportions for the “Colours” domain. Analogous figures for all domains are in the Supplementary Materials.’



simplify to  $\hat{P}_T(a) \geq \hat{P}_{\{a,b\}}(a)\hat{P}_{\{a,c\}}(a)$ ,  $\hat{P}_T(b) \geq \hat{P}_{\{b,a\}}(b)\hat{P}_{\{b,c\}}(b)$  and  $\hat{P}_T(c) \geq \hat{P}_{\{c,a\}}(c)\hat{P}_{\{c,b\}}(c)$ . The vector  $(\hat{P}_T(a), \hat{P}_T(b), \hat{P}_T(c))$  of ternary choice proportions lies outside the red triangle and in particular,  $\hat{P}_T(b) = 0.375$  is less than  $\hat{P}_{\{a,b\}}(b)\hat{P}_{\{b,c\}}(b) = 0.475 \cdot 0.925 = 0.438375$ .

#### 5.4 Tests of axioms using Bayes factors

- 5 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.

Kass & Raftery (1995) classify log Bayes factors between 0 and 1 as “not worth more than a  
10 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.

Figure 4 gives, for each choice domain, log Bayes factors for various combinations of axiom and encompassing model, relative to encompassing model  $M_3$ . There is a panel for each domain, with  
15 three columns of points, corresponding to the three encompassing models  $M_1$ ,  $M_2$  and  $M_3$ . 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  $\Lambda_{\text{Reg}}$ ,  $\Lambda_{\text{RU}}$  and  $\Lambda_{\text{MI}}$ , respectively. 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 relative

Menu $A$	Choice counts					Choice proportions				
	$N_A(a)$	$N_A(b)$	$N_A(c)$	$N_A(d)$	$N_A(e)$	$\hat{P}_A(a)$	$\hat{P}_A(b)$	$\hat{P}_A(c)$	$\hat{P}_A(d)$	$\hat{P}_A(e)$
$\{a, b\}$	21	19	-	-	-	0.525	0.475	-	-	-
$\{a, c\}$	28	-	12	-	-	0.700	-	0.300	-	-
$\{b, c\}$	-	37	3	-	-	-	0.925	0.075	-	-
$\{a, b, c\}$	19	15	6	-	-	0.475	0.375	0.150	-	-
$\{a, d\}$	8	-	-	32	-	0.200	-	-	0.800	-
$\{b, d\}$	-	15	-	25	-	-	0.375	-	0.625	-
$\{a, b, d\}$	7	8	-	25	-	0.175	0.200	-	0.625	-
$\{c, d\}$	-	-	6	34	-	-	-	0.150	0.850	-
$\{a, c, d\}$	10	-	3	27	-	0.250	-	0.075	0.675	-
$\{b, c, d\}$	-	5	7	29	-	-	0.122	0.171	0.707	-
$\{a, b, c, d\}$	6	4	3	27	-	0.150	0.100	0.075	0.675	-
$\{a, e\}$	22	-	-	-	18	0.550	-	-	-	0.450
$\{b, e\}$	-	16	-	-	24	-	0.400	-	-	0.600
$\{a, b, e\}$	16	12	-	-	12	0.400	0.300	-	-	0.300
$\{c, e\}$	-	-	11	-	29	-	-	0.275	-	0.725
$\{a, c, e\}$	12	-	9	-	19	0.300	-	0.225	-	0.475
$\{b, c, e\}$	-	16	8	-	16	-	0.400	0.200	-	0.400
$\{a, b, c, e\}$	14	11	3	-	12	0.350	0.275	0.075	-	0.300
$\{d, e\}$	-	-	-	33	7	-	-	-	0.825	0.175
$\{a, d, e\}$	6	-	-	31	3	0.150	-	-	0.775	0.075
$\{b, d, e\}$	-	10	-	20	10	-	0.250	-	0.500	0.250
$\{a, b, d, e\}$	4	9	-	24	4	0.098	0.220	-	0.585	0.098
$\{c, d, e\}$	-	-	4	30	6	-	-	0.100	0.750	0.150
$\{a, c, d, e\}$	7	-	1	21	11	0.175	-	0.025	0.525	0.275
$\{b, c, d, e\}$	-	11	2	21	6	-	0.275	0.050	0.525	0.150
$\{a, b, c, d, e\}$	3	11	3	17	6	0.075	0.275	0.075	0.425	0.150

Table 4: Observed choice counts and proportions 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. Analogous tables for all domains are in the Supplementary Materials.

to itself is zero. 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 in support of 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 ?? and 4, 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  $\lambda$ .

## 5.5 Posterior distributions of $\alpha$ and $\lambda$

Recall that the concentration parameter  $\alpha$  governs how consistent choices are likely to be from trial to trial. As  $\alpha$  approaches zero, choice probabilities near zero or one become more probable, and choice becomes more consistent from trial to trial; as  $\alpha$  approaches infinity, choice probabilities near the discrete uniform (where objects in a menu are equally likely to be chosen) become more probable and choice becomes more variable across trials. We would expect  $\alpha$  to vary considerably across domains. Not surprisingly, the posterior distribution of  $\alpha$  varies from domain to domain. The domain with the least posterior median was “Surface area”, where the posterior median was 5.6 and the 90% credible set<sup>2</sup> was [3.1, 9.4]. The domain with the greatest posterior median was “Colour Combos”, with a posterior median of 17.4 and a credible set of [9.1, 30.5].

Values of the posterior median of  $\alpha$  are much higher than those reported in McCausland &

---

<sup>2</sup>computed as the interval from the 5th percentile to the 95th

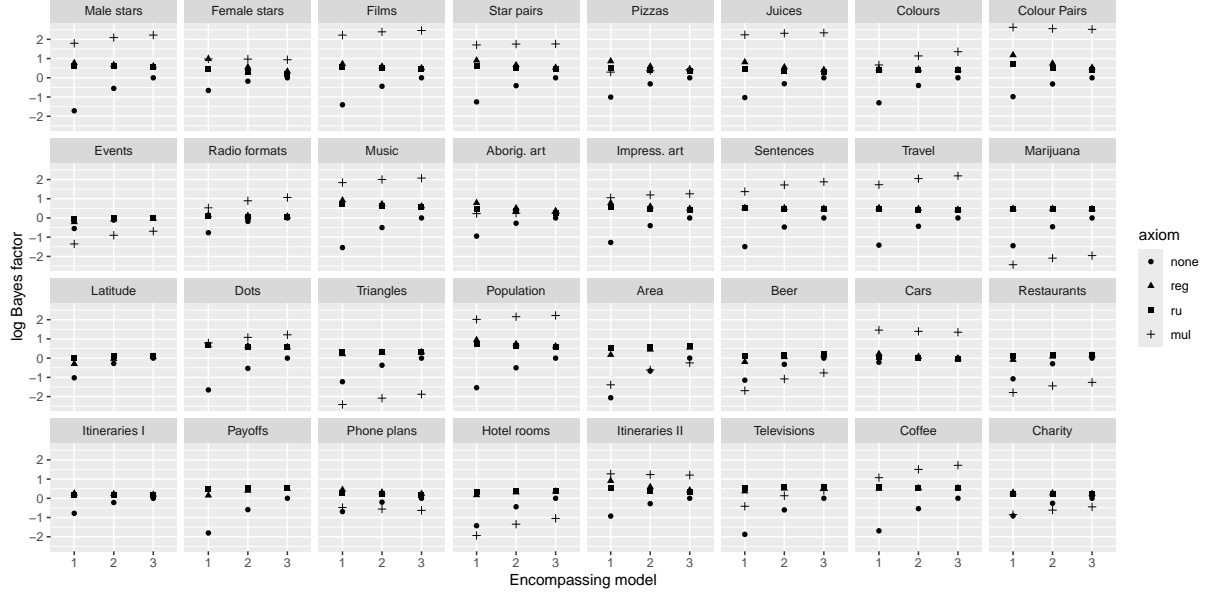


Figure 4: Log Bayes factors for 32 domains, by encompassing model and axiom. All log Bayes factors are relative to encompassing model  $M_3$ . The “encompassing model” indices 1, 2 and 3 correspond to models  $M_1$ ,  $M_2$  and  $M_3$

Marley (2014) and McCausland et al. (2020), which analysed data from experiments with a within-subject design. Again, this is not surprising. We would expect the repeated choices of a single individual to be more consistent than the choices of a sample of individuals.

Table ?? shows numerical estimates of quantiles of  $\lambda$ . 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 ?? except that the quantiles for  $p = 0.05$ ,  $p = 0.25$  and  $p = 0.5$  are tabulated. The posterior quantiles of  $\lambda$  for  $p = 0.75$  and  $p = 0.95$  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) 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. 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 4 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).

## 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 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 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 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 complement the models introduced in McCausland & Marley (2013) and the inferential methods described in McCausland & Marley (2014). Together, the models, methods and data serve as a testing ground for the axioms we have considered in this 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 in support of 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  $\lambda$ ) similar to the kind of dependence associated with random utility, and that encompassing models with a higher prior mean of  $\lambda$  are favoured. Evidence in support of 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 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 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  
5 with fewer attributes.

The section titles, “Male stars” for example, give the names of the domains for the purpose of 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  
10 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 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,  
15 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.

## A.1 No numerical attributes

### A.1.1 Male stars

The source for this domain is the website `ranker.com`, accessed June 4, 2017. The list is “The best  
20 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

### A.1.2 Female stars

The source for this domain is the website [ranker.com](http://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

### A.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.

#### A.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

#### A.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

#### A.1.6 Juices

Which one of the following fresh juices would you choose?

- Mango
- Orange
- Apple
- Grapefruit
- Pineapple

### A.1.7 Colours

Which one of the following colours do you like best?

- Red
- Purple
- Pink
- Blue
- Green

### A.1.8 Colour combos

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

### A.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.

#### A.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 A.1.11 Music

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

#### A.1.12 Aboriginal

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



#### A.1.13 Impressionist

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



5

#### A.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.

10



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?

#### **A.1.15 Travel**

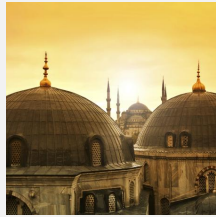
The source is Tripadvisor. These are the top five travel destinations, according to the results of an on-line contest where visitors to a Tripadvisor site could make pairwise choices between travel

5 destinations.

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



1. Marrakech, Morocco



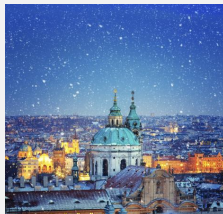
2. Istanbul, Turkey



3. Hanoi, Vietnam



4. Siem Reap, Cambodia



5. Praque, Czech Republic

### A.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.

## A.2 Single attributes, not directly observed

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

### A.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.

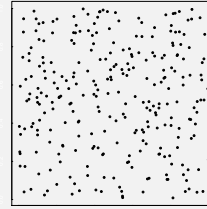
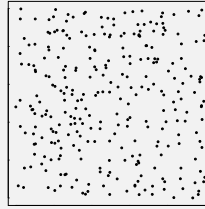
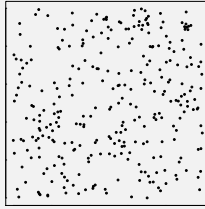
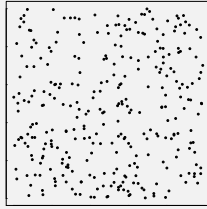
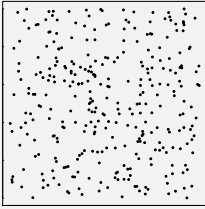
Which one of the following cities do you think is furthest north?

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

### A.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.

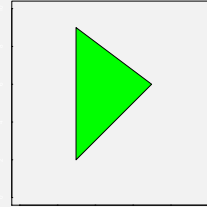
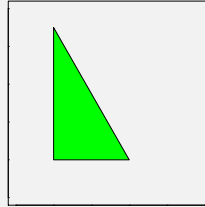
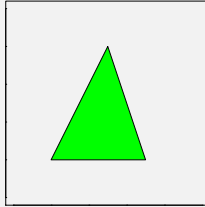
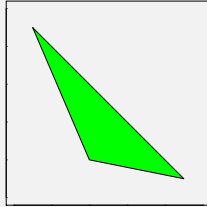
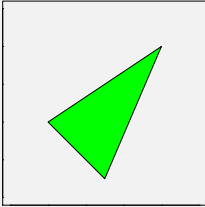
Which one of the following boxes do you think has the greatest number of points?



### A.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

### A.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.

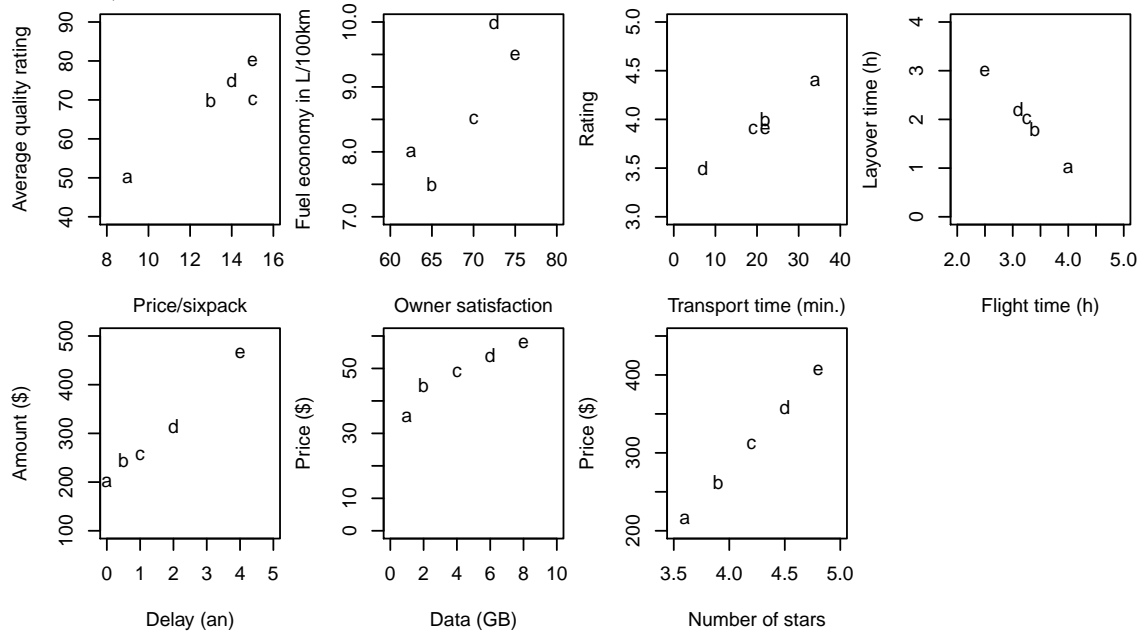
Which one of the following countries do you think has the largest population?

- Indonesia
- Brazil
- Pakistan
- Nigeria
- Bangladesh

### 10 A.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 5: 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

### A.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 5 plots the five objects of each of these domains in attribute space, revealing potential conditions for context effects.

Table 5 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 5: 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.

### A.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 5 show the choice objects in attribute space. Table 5 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

### A.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 5 shows

the choice objects in attribute space. Table 5 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

### A.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 5 shows the choice objects in attribute space. Table 5 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

### A.3.4 Layovers

10 The fourth panel of Figure 5 shows the choice objects in attribute space. Table 5 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

### A.3.5 Future

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.

### A.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.

### A.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

## A.4 Objects with multiple attributes

### A.4.1 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

#### A.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 A.4.3 Coffee

The source for this domain is the website [buycoffeecanada.com](http://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

#### A.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

### Supplementary Materials

Supplementary Materials can be found in the folder `Supplementary_materials` of the Github repository <https://github.com/mccauslw/Population>.

Figures analogous to Figure 3 and tables analogous to Table 4 for each of the 32 domains are in the document `all_domains_data.pdf`. 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  
5 `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`.

## Authors' Contribution Statement

McCausland contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, project administration, software, visualization and writing (initial draft, review and  
10 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

The authors thank the Institute for Choice at the University of South Australia for programming  
15 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

The authors declare none.

## 20 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

Bard, E. G., Robertson, D. & Sorace, A. (1996), 'Magnitude estimation of linguistic acceptability',  
25 *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-Verlag, New York, NY.
- 5 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 Probability and Statistics: Essays in Honor of Harold Hotelling’, Stanford University Press,  
10 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.
- 15 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  
20 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.
- 25 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.
- 5 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 the random utility hypothesis directly’, *The Economic Journal* **130**, 183–207.
- 10 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.
- 15 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.
- 20 Simonson, I. (1989), ‘Choice based on reasons: The case of attraction and compromise effects’, *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.
- 25 Tversky, A. (1972b), ‘Elimination by aspects: A theory of choice’, *Psychological Review* **79**(4), 281–299.

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.

- 5 Zwillling, 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.