BAYESIAN INFERENCE AND MODEL COMPARISON FOR RANDOM CHOICE STRUCTURES

WILLIAM J. MCCAUSLAND, SCIENCES ÉCONOMIQUES AND CIREQ, UNIVERSITÉ DE MONTRÉAL AND A. A. J. MARLEY, DEPARTMENT OF PSYCHOLOGY, UNIVERSITY OF VICTORIA

Abstract. We complete the development of a testing ground for axioms of discrete stochastic choice. Our contribution here is to develop new posterior simulation methods for Bayesian inference, suitable for a class of prior distributions introduced by McCausland and Marley (2013). These prior distributions are joint distributions over various choice distributions over choice sets of different sizes. Since choice distributions over different choice sets can be mutually dependent, previous methods relying on conjugate prior distributions do not apply. We demonstrate by analyzing data from a previously reported experiment and report evidence for and against various axioms.

Key words and phrases. Random utility, discrete choice, Bayesian inference, MCMC.
Financial support from the SSHRC is gratefully acknowledged. JEL codes: C11, C35, C53, D01.

1. Introduction

We consider an environment where agents face various choice sets A, all subsets of the same finite master set $T = \{x_1, \ldots, x_n\}$ of objects. Agents choose a single object from a choice set A each time it is presented to them.

Most models for stochastic discrete choice specify or imply choice probabilities $P_A(x)$, for all $x \in A \subseteq T$. We assume that these choice probabilities describe the choice behaviour of a single agent. This assumption holds for the data we analyse here; alternatively, we could interpret choice probabilities as describing the choice behaviour of agents randomly drawn from some population. We also assume that choices are statistically independent across presentations of choice sets.

A *random choice structure* (T, P) is the complete specification of the $P_A(x)$. As such, a random choice structure with no restrictions on probabilities is a non-parametric model. It is true that it consists of a finite number of unknown probabilities, but this is a consequence of the finite nature of choice sets, not the imposition of a restrictive finite-dimensional parametric distribution.

With flexibility comes the danger of over-fitting and poor out-of-sample predictive performance. Prior information can impose discipline, and it can come in many forms, including choice axioms imposing constraints on probabilities across choice sets. Various axioms have been suggested in the literature. See below for some examples and McCausland and Marley (2013) for further discussion, including graphical illustrations of the relationships among them and citations to the literature.

The purpose of this paper is to propose, implement and demonstrate a testing ground for probabilistic choice axioms in an abstract choice setting. It involves applying methods of Bayesian model comparison to measure the plausibility of axioms in the light of discrete choice data. These include compound axioms, obtained as the union, intersection or complement of other axioms. We investigate several particular axioms, but emphasize that our approach can be used to evaluate others, including those yet to be proposed.

1.1. Some axioms from the literature.

Some axioms pertain only to binary choice probabilities. Due to the importance of these probabilities, we adopt a standard notational convention: for all distinct $x, y \in T$, we write $p(x, y)$ for $P_{\{x,y\}}(x)$. The random choice structure (T, P) satisfies

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

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

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

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

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

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

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

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

Other axioms constrain choice probabilities on differently sized choice sets. We say that (T, P) satisfies

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

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

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

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

For MI, see Sattath and Tversky (1976), Colonius (1983) and Suck (2002). For the remaining conditions, see Luce and Suppes (1965). MI should not be confused with the multiplication *condition* in Luce and Suppes (1965), which is a different axiom, involving only binary choice probabilities. McCausland and Marley (2013) survey in more detail the literature on theorems about these axioms, and graphically illustrate some of the relationships among them.

We will need some more notation to define a final condition. For all non-empty $A \subseteq T$, we define $R(A)$ as the set of rankings on A; a *ranking distribution* on A is a pair (A, Π) such that Π is a probability mass function on $R(A)$. For any ranking distribution (T, Π), we define the random choice structure *induced* by (T, Π) as the random choice structure (T, P^{Π}) such that for all non-empty $A \subseteq T$, and all $x \in A$,

$$P^{\Pi}_A(x) = \sum_{\{\succ \in R(T)\colon\, h_{\succ}(A)=x\}} \Pi(\succ),$$

where for every nonempty $A \subseteq T$ and every rank order $\succ \in R(T)$, $h_{\succ}(A)$ is the highest $\succ$-ranked object in A.

Our final condition is this: a random choice structure (T, P) satisfies the *random ranking hypothesis*, denoted RR, if there is a ranking distribution (T, Π) such that $P = P^{\Pi}$. While this definition is not framed in terms of choice probabilities, there are necessary and sufficient conditions that are. Fiorini (2004) gives these conditions as follows: for all non-empty $A \subseteq T$ and all $x \in A$,

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

Block and Marschak (1960) and Luce and Suppes (1965, Theorem 49) show that for finite master sets the random ranking hypothesis is equivalent to what is often known as "random utility". Random utility models are those in which agents select from each choice set as if they drew, independently and from the same continuous distribution, a random utility function over the master set and then went on to choose the utility maximizing element from that set. The assumption that utilities have a continuous distribution implies that the probability that any two utilities are equal is zero. If the definition is only asserted for the binary choice probabilities, then the model is called a *binary* random utility model. If the utilities u_x, $x \in T$, are mutually independent, then we say the model is an *independent* random utility model. When the master set has no more than five elements, TI is necessary and sufficient for binary random utility. See Dridi (1980) for a proof, Koppen (1995) and the literature cited there for additional necessary conditions when the master set has more than five elements. Sattath and Tversky (1976) show that MI is necessary for an independent random utility model.

There is a relatively long history in Economics, Psychology and Marketing, of theory and application of probabilistic discrete choice models. Most of these models are random utility models. Widely used random utility models include the (multinomial) logit, (multinomial) probit, McFadden's (1977) Generalized Extreme Value (GEV) model, the class of mixed (multinomial) logit models and Tversky's (1972) Elimination By Aspects (EBA) model.

Logit models are independent random utility models by construction. Probit models are random utility models, also by construction, but not necessarily independent random utility models. The class of GEV models explicitly includes logit, nested logit, paired combinatorial logit and generalized nested logit models. McFadden (1977) shows that a representation of choice probabilities characterizing GEV is equivalent to a random utility model where the vector of utilities has a generalized extreme value distribution. Dagsvik (1994) shows that the GEV class is dense in the set of random utility models. The class of mixed logit models explicitly includes latent class logit models. McFadden and Train (2000) show a limiting equivalence of the set of mixed multinomial logit models and the set of random utility models. See Train (2009) for more on logit, probit, GEV and mixed logit.

The EBA model is not explicitly constructed as a random utility model, but Tversky (1972, Theorem 7) shows that it is indeed one. Sattath and Tversky (1976) show that EBA models satisfy MI, which we have seen is a necessary condition for independent random utility; however, Tversky (1972) gives an example of an EBA model that is not an independent random utility model.

In Economics and Marketing, probabilistic discrete choice models are almost exclusively random utility ones. In Psychology, random utility models, including logit, probit and EBA, are commonly used. See summaries in Luce and Suppes (1965), Luce (1977), Luce (1994) and Marley's (1992a, 1992b, 2002) editorial introductions to special journal issues. Models that are not necessarily random utility models include dynamic stochastic choice models such as *decision field theory* models and the *leaky competing accumulator model.* These are summarized in Rieskamp, Busemeyer, and Mellers (2006) and Busemeyer and Rieskamp (2013).

1.2. **Statistical methods for testing axioms.** There is a long history of using data on observed choice frequencies to support or undermine probabilistic choice axioms. Regenwetter, Dana, and Davis-Stober (2011) survey some of the approaches used in the literature on stochastic transitivities. Many studies interpret frequencies as probabilities, and measure the evidence for or against an axiom by the number of necessary conditions that are violated; such an approach ignores sampling variation. Other studies take into account sampling variability, but run into multiple testing problems, by performing multiple tests of various necessary conditions rather than a single joint test of a set of necessary and sufficient conditions. Another issue is using distributions for test statistics that are not even asymptotically correct under the null hypothesis that an axiom holds; correct frequentist inference is notoriously difficult when parameter values are subject to inequality constraints and point estimates of parameters are near the boundary of the constrained set. The above problems can well lead to erroneous conclusions; in addressing them, Iverson and Falmagne (1985) overturn the conclusions of Tversky (1969).

Cavagnaro and Davis-Stober (2013), Myung, Karabatsos, and Iverson (2005) and Zwilling, Cavagnaro, and Regenwetter (2011) take a Bayesian approach to testing axioms taking the form of inequality restrictions over probabilities. Testing these constraints or estimating parameters subject to them is conceptually straightforward in a Bayesian framework. A baseline model, consisting of a prior distribution over the set of relevant choice probabilities, serves as an encompassing model. A restricted model is obtained by truncating the prior distribution to the set of probability configurations that satisfy some axiom. The Bayes factor in favour of the restricted model against the baseline model is equal to the ratio of posterior to prior probability of the restriction holding in the baseline model.

Cavagnaro and Davis-Stober (2013) and Myung, Karabatsos, and Iverson (2005) both use a uniform prior on the space of relevant binary choice probabilities to define their baseline model. Probabilities for distinct pairs of objects are independent and their marginal distributions are all uniform on $[0, 1]$. Truncation to the region where some axiom holds typically induces dependence and non-uniform marginals. Myung, Karabatsos, and Iverson (2005) discuss two possible extensions, to non-uniform priors and non-binary probabilities. They suggest Beta distributions as non-uniform priors for binary choice probabilities and Dirichlet priors for non-binary choice probabilities. Their claim that these priors are conjugate for a likelihood function arising from choice observations implies that they have in mind a joint prior distribution where choice probabilities over distinct choice sets are independent.

McCausland and Marley (2013) introduces a family of joint distributions over all the choice probabilities in a random choice structure. The marginal distributions are symmetric Dirichlet, but choice probabilities across choice sets need not be independent. As far as we know, this is the first paper to propose a baseline model where choice probabilities are dependent. Unfortunately, this dependence destroys conjugacy, which makes it more difficult to simulate from the posterior distribution. Until now, these priors have not been used for empirical analysis. The present paper develops the posterior simulation methods needed for inference.

1.3. **Empirical evidence for and against various axioms.** Rieskamp, Busemeyer, and Mellers (2006) review the empirical literature testing weak and strong stochastic transitivity and regularity. They conclude that although some have found systematic violations of weak stochastic transitivity, the violations are limited to rare and unusual situations. However, they point to an "overwhelming number of studies" suggesting that human behaviour does not satisfy strong stochastic transitivity.

They also document much evidence against the regularity axiom. Since regularity is necessary for random utility, violations of the former are violations of the latter. They identify different types of regularity violations, including attraction and asymmetrical dominance effects.

To our knowledge, the multiplicative inequality has not been tested directly. Independent random utility, a stronger condition, is considered by many to be too inflexible, but it is not known how consistent the multiplicative inequality is with observed choices.

1.4. **Prior distributions for random choice structures.** Bayesian analysis involves the choice of a prior distribution. McCausland and Marley (2013) propose a class of prior distribution on the space of random choice structures, indexed by two parameters, α and λ. The α parameter governs how consistent an agent is likely to be in repeated choices from the same choice set; for low values of α, a random choice structure drawn from the prior is likely to feature choice probabilities $P_A(x)$ close to zero and one; for high values of α, they are likely all to be close to $1/|A|$. The λ parameter governs the degree of dependence of choice probabilities across choice sets. For $\lambda = 0$, the vectors $(P_A(x))_{x\in A}$ are mutually independent, $A \subseteq T$; thus learning $P_A(\cdot)$ gives no information about $P_B(\cdot)$. For $\lambda = 1$, the random choice structure satisfies the random ranking hypothesis with probability one. While we do not know the joint density over the space of random choice structures in closed form, we do know the marginal distributions. They are

$$(P_A(x_1), \ldots P_A(x_{|A|})) \sim \mathrm{Di}\left(\frac{\alpha}{|A|!}, \ldots, \frac{\alpha}{|A|!}\right),$$

where $\mathrm{Di}(\cdot)$ denotes the Dirichlet distribution — see Forbes, Evans, Hastings, and Peacock (2011).

1.5. **Outline.** Section 2 describes a model for discrete stochastic choice, consisting of a hierarchical prior distribution for a random choice structure (T, P). The highest level of the hierarchy gives a prior distribution for the hyper-parameters α and λ of the class of priors in McCausland and Marley (2013).

Section 3 describes Bayes factors, which we use to document the evidence for or against various axioms of discrete stochastic choice. In all the cases we consider, the event that an axiom holds has non-zero prior probability. In these cases, the Bayes factor of an axiom, with respect to a baseline model, equals the ratio of posterior to prior probabilities of the axiom holding in the baseline model.

Section 4 describes posterior simulation methods. It does not help us that the marginal prior distribution of each $P_A(\cdot)$ is Dirichlet, the conjugate distribution for the likelihood function for independent categorical data. A consequence of our decision to allow prior dependence across choice sets is that the joint prior distribution over all choice probabilities is not conjugate for the entire likelihood function. For this reason, we resort to Markov chain Monte Carlo (MCMC) simulation methods to simulate from the posterior distribution and thereby compute posterior moments and quantiles of interest.

Section 5 reports results from the analysis of data from previous experiments. Section 6 concludes.

2. An Unrestricted Model for Discrete Stochastic Choice

A random choice structure (T, P) gives a family of distributions for discrete stochastic choice. Here we complete the model by specifying a hierarchical prior distribution for the random choice structure (T, P). We will call this the *unrestricted* model and denote it M_0. We will also consider various restricted models, obtained by imposing different choice axioms.

The prior specifies the joint distribution of two hyper-parameters δ and $\tilde{\delta}$, a vector γ of latent variables and the random choice structure (T, P). At the upper level of the hierarchy are two hyper-parameters, δ and $\tilde{\delta}$, *a priori* independent with distributions

$$\delta \sim \mathrm{Ga}(a, b), \qquad \tilde{\delta} \sim \mathrm{Ga}(\tilde{a}, \tilde{b}). \tag{2}$$

The two parameters α and λ in McCausland and Marley (2013) are given as the following transformations of δ and $\tilde{\delta}$:

$$\lambda = \frac{\delta}{\delta + \tilde{\delta}}, \qquad \alpha = \delta + \tilde{\delta}.$$

We use δ and $\tilde{\delta}$ only for computational convenience; α and λ are the parameters of interest. When $b = \tilde{b}$, the implied joint prior distribution of α and λ is such that α and λ are independent, with

$$\lambda \sim \mathrm{Be}(a, \tilde{a}), \quad \alpha \sim \mathrm{Ga}(a + \tilde{a}, b).$$

The next level of the hierarchy gives the conditional distribution of latent variables given hyper-parameters, a distribution described in McCausland and Marley (2013). Given hyper-parameters, the latent variables are conditionally independent. For each ranking $\succ \in R(T)$, there is a latent variable $\gamma(\succ)$ with conditional distribution

$$\gamma(\succ) | \delta, \tilde{\delta} \sim \mathrm{Ga}\left(\frac{\delta}{n!}, 1\right). \tag{3}$$

For each choice set A and each ranking $\succ \in R(A)$, there is a latent variable $\tilde{\gamma}_A(\succ)$ with conditional distribution

$$\tilde{\gamma}_A(\succ) | \delta, \tilde{\delta} \sim \mathrm{Ga}\left(\frac{\tilde{\delta}}{|A|!}, 1\right). \tag{4}$$

The lowest level of the hierarchy gives choice probabilities as deterministic functions of the latent variables:

$$P_A(x) = \frac{\sum_{\succ \in R(T):\, x = h_\succ(A)} \gamma(\succ) + \sum_{\succ \in R(A):\, x = h_\succ(A)} \tilde{\gamma}_A(\succ)}{\sum_{\succ \in R(T)} \gamma(\succ) + \sum_{\succ \in R(A)} \tilde{\gamma}_A(\succ)}. \tag{5}$$

We denote by γ the vector of all weights $\gamma(\succ)$ and $\tilde{\gamma}_A(\succ)$.

We use the same prior distribution for all participants in all experiments, and do posterior inference for each participant separately. Alternatively, one could extend the hierarchical prior to induce dependence of random choice structures across participants — the resulting joint analysis would "borrow strength" across individuals — but we do not pursue this here.

Thus, we do not need to introduce notation to distinguish participants in the experiment. For the remainder of the section, we assume we are discussing the choices of a single participant.

For every $A \subseteq T$ and $x \in A$, we observe $N_A(x)$, the number of times the participant chooses object x when presented with choice set A. For all $A \subseteq T$, let N_A be the vector $(N_A(x))_{x\in A}$ of all choice counts associated with A. Let N be the vector of all choice counts, $(N_A(x))_{A\subseteq T, x\in A}$. In some cases, there will be a choice set B the participant never sees. In such a case, the vector $N_B(\cdot)$ will be zero. However, since the $P_A(\cdot)$, $A \subseteq T$, are statistically dependent across choice sets, the posterior distribution of $P_B(\cdot)$ will typically not be the same as its prior distribution.

Since we assume choice events are independent across trials, the log likelihood function can be written as

$$\mathcal{L}(\gamma; N) = \sum_{A\subseteq T} \sum_{x\in A} N_A(x) \log P_A(x).$$

It will be helpful to decompose the log likelihood by choice set. Accordingly, we write

$$\mathcal{L}(\gamma; N) = \sum_{A\subseteq T} \mathcal{L}_A(\gamma; N), \qquad \text{where} \qquad \mathcal{L}_A(\gamma; N) = \sum_{x\in A} N_A(x) \log P_A(x).$$

3. Bayes Factors

We evaluate the plausibility of an axiom in the light of observed data by reporting a simulation consistent approximation of the Bayes factor in favour of a restricted model M_r, in which the axiom holds, against the unrestricted model M. By Bayes' rule, we can express this Bayes factor as

$$\frac{\Pr[N|M_r]}{\Pr[N|M]} = \frac{\Pr[\Lambda|N, M]}{\Pr[\Lambda|M]},$$

where Λ is the event that the axiom holds for (T, P).

The left hand side gives the Bayes factor as it is usually defined, in terms of a ratio of marginal likelihoods. The right hand side is a ratio of the posterior to the prior probability of the axiom holding in the unrestricted model. A high posterior probability is a measure of how consistent the data are with the axiom; a low prior probability is a measure of how small or parsimonious the model becomes when the axiom is imposed. In McCausland and Marley (2013), we pointed out that since the numerator probability cannot exceed one, the reciprocal of an axiom's prior probability gives an upper bound on the Bayes factor in favour of the restricted model in which the axiom holds. No matter how much data is collected for a single decision maker, the Bayes factor cannot exceed this bound.

We will approximate the numerator and denominator probabilities using prior and posterior simulation, respectively, and compute numerical standard errors measuring simulation noise.

4. Prior and Posterior Simulation

Most techniques of Bayesian empirical analysis involve computing moments and quantiles of prior or posterior distributions of unknown quantities. Prime examples include point and interval estimation, model comparison, prior and posterior predictive analysis, and out-of-sample prediction. See Berger (1985), Bernardo and Smith (1994) and Geweke (2005). In our case, we will be computing prior and posterior probabilities, which are means of indicator functions, as well as prior and posterior moments of the α and λ parameters.

Closed form evaluation of many prior and most posterior moments and quantiles is intractable, so practitioners usually resort to Monte Carlo simulation methods. First, they draw a sample from the appropriate target distribution; then they approximate moments and quantiles of the target by their sample counterparts. Independence Monte Carlo, based on an iid sample, is usually practical when the target is the prior distribution but not when it is the posterior. For the posterior distribution, most resort to Markov chain Monte Carlo methods. Laws of large numbers and central limit theorems for ergodic Markov chains are available to describe and measure simulation error. For texts introducing MCMC, see Gilks, Richardson, and Spiegelhalter (1996) and Robert and Casella (2010). For details on basic Markov chain asymptotic theory, see Meyn and Tweedie (1993).

We will report posterior moments of α and λ, and Bayes factors in favor of various axioms, for six different baseline models, M_1 through M_6, specified in Section 5. The six models differ in terms of the prior, and the purpose of multiple models is to illustrate the sensitivity of results to the prior specification.

While we are only interested in results for the six baseline models, we simulate from the prior and posterior distributions of a different model, M_0. We then use importance sampling to compute prior and posterior probabilities and other moments for the six baseline models. We never simulate directly from the baseline models. The prior of M_0 is such that all importance weights are bounded. See Geweke (1989) for more on importance sampling.

Prior simulation is straightforward: we obtain an i.i.d. sample by direct simulation from the gamma distributions in (2), (3) and (4). We use routines from the GNU Scientific Library to draw gamma random variables.

Posterior simulation is more difficult, and we resort to MCMC methods. In Section 4.1 and Appendix A, we describe the Markov chains we use to sample from the posterior distribution.

In Section 4.1.4 we describe how to use the prior and posterior samples we obtain for model M_0 to compute prior and posterior moments for the six baseline models $M_1, \ldots, M_6$. We use importance sampling, which amounts to re-weighting the various draws from the posterior sample such that weighted sample moments approximate population moments for one of the six baseline models. We also show how to compute numerical standard errors, a measure of simulation noise.

Computing prior and posterior probabilities of axioms involves repeated evaluation of an indicator function over several different random choice structures. To determine whether an axiom holds for a given random choice structure, we use the robust methods described in McCausland and Marley (2013), to guard against classification errors due to machine rounding error.

4.1. **Posterior simulation.** We now describe an ergodic Markov chain whose invariant distribution is the posterior distribution for the unrestricted model in Section 2. The posterior distribution is the conditional distribution of hyper-parameters δ, $\tilde{\delta}$ and γ given data N. As in many chains used for posterior simulation, the random transition from the current state of the chain to the next consists of a sequence of several Metropolis-Hastings transitions, each updating some of the unknown quantities of the model in such a way as to preserve the posterior distribution. When we say that a stochastic transition preserves a distribution we mean that the distribution is an invariant distribution of the transition. See Chib and Greenberg (1995) for a tutorial on the Metropolis-Hastings algorithm.

A single transition of the chain consists of a sequence of three Metropolis-Hastings updates, described in Sections 4.1.1, 4.1.2 and 4.1.3. Once we have a posterior sample $\gamma^{(j)}$, $j = 1, \ldots, J$, we can obtain a posterior sample $P^{(j)}$, $j = 1, \ldots, J$, using (5), draw by draw.

4.1.1. *A Metropolis-Hastings update for δ and $\gamma(\succ)$, $\succ \in R(T)$.* The first update is a Metropolis-Hastings transition replacing current values δ and $\gamma(\succ)$, $\succ \in R(T)$, with random new values δ' and $\gamma'(\succ)$, $\succ \in R(T)$. It preserves the conditional distribution of δ and $\gamma(\succ)$, $\succ \in R(T)$, given $\tilde{\delta}$, other latent variables, and data N.

(1) Draw $\beta \sim \text{Be}(\pi a, (1-\pi)a)$ and $\epsilon \sim \text{Ga}((1-\pi)a, b)$, with β and ϵ mutually independent and independent of the history of the chain, and form the candidate value $\delta^* = \beta\delta + \epsilon$. Since β and ϵ are only devices used to obtain δ^*, they are discarded. The random transition from δ to δ^* is an example of a Beta-Gamma transition, and it preserves the conditional distribution of δ given a and b — see Appendix A. Here, $\pi \in (0, 1)$ is a fixed parameter governing the degree of dependence between δ and δ^*.

(2) For all $\succ \in R(T)$,
   (a) if $\delta^* > \delta$, draw the proposal $\gamma^*(\succ)$ from the following conditional distribution of $\gamma^*(\succ)$ given $\gamma(\succ)$, δ and δ^*:
$$\gamma^*(\succ) - \gamma(\succ) \sim \text{Ga}\left(\frac{\delta^* - \delta}{n!}, 1\right).$$
   (b) if $\delta^* \leq \delta$, draw $\gamma^*(\succ)$ from the following conditional distribution:
$$\frac{\gamma^*(\succ)}{\gamma(\succ)} \sim \text{Be}\left(\frac{\delta}{n!}, \frac{\delta - \delta^*}{n!}\right).$$

(3) Jointly accept the proposal consisting of δ^* and $\gamma^*(\succ)$, $\succ \in R(T)$, with probability
$$\min\left(\frac{\mathcal{L}(\gamma^*; N)}{\mathcal{L}(\gamma; N)}, 1\right).$$
Accepting the proposal means setting new values equal to proposals; here, setting $\delta' = \delta^*$ and $\gamma'(\succ) = \gamma^*(\succ)$, $\succ \in R$. Rejecting means setting new values equal to old values; here, setting $\delta' = \delta$ and $\gamma'(\succ) = \gamma(\succ)$, $\succ \in R$.

Appendix A shows that the update described here is a true Metropolis-Hastings update of the conditional distribution of δ and $\gamma(\succ)$, $\succ \in R$, given data, other parameters and other latent variables.

4.1.2. *A Metropolis-Hastings update for* $\tilde{\delta}$ *and* $\tilde{\gamma}_A(\succ)$, $A \subseteq T$, $\succ \in R(A)$. The second update does something very similar for the hyper-parameter $\tilde{\delta}$ and the $\tilde{\gamma}_A(\succ)$, $A \subseteq T$ and $\succ \in R(A)$.

(1) Draw $\beta \sim \mathrm{Be}(\pi\tilde{a}, (1-\pi)\tilde{a})$ and $\epsilon \sim \mathrm{Ga}((1-\pi)\tilde{a}, \tilde{b})$, independently, and form $\tilde{\delta}^* = \beta\tilde{\delta} + \epsilon$.
(2) For all non-empty $A \subseteq T$ and $\succ \in R(A)$,
  (a) if $\tilde{\delta}^* > \tilde{\delta}$, draw
$$\tilde{\gamma}^*_A(\succ) - \tilde{\gamma}_A(\succ) \sim \mathrm{Ga}\left(\frac{\tilde{\delta}^* - \tilde{\delta}}{|A|!}, 1\right)$$
  (b) if $\tilde{\delta}^* \leq \tilde{\delta}$, draw
$$\frac{\tilde{\gamma}^*_A(\succ)}{\tilde{\gamma}_A(\succ)} \sim \mathrm{Be}\left(\frac{\tilde{\delta}}{|A|!}, \frac{\tilde{\delta} - \tilde{\delta}^*}{|A|!}\right)$$
(3) Jointly accept $\tilde{\delta}^*$ and $\tilde{\gamma}^*_A(\succ)$, $A \subseteq T$ and $\succ \in R(A)$, with probability
$$\min\left(\frac{\mathcal{L}(\gamma^*; N)}{\mathcal{L}(\gamma; N)}, 1\right).$$

Appendix A shows that this update is a true Metropolis-Hastings update of the conditional distribution of $\tilde{\delta}$ and $\tilde{\gamma}_A(\succ)$, $A \subseteq T$ and $\succ \in R$, given data, other parameters and other latent variables.

4.1.3. *A Metropolis-Hastings update for* $\tilde{\gamma}_A$.

(1) For all $A \subseteq T$ and $\succ \in R(A)$,
  (a) draw $\tilde{\gamma}^*_A(\succ) \sim \mathrm{Ga}\left(\frac{\tilde{\delta}}{|A|!}, 1\right)$,
  (b) accept $\tilde{\gamma}^*_A(\succ)$ with probability
$$\min\left(\frac{\mathcal{L}_A(\gamma^*; N)}{\mathcal{L}_A(\gamma; N)}, 1\right).$$

This a sequence of direct Metropolis updates, each updating the conditional distribution of one of the $\tilde{\gamma}^*_A(\succ)$ given everything else. These updates do not change the state of the chain by much. Furthermore, they are redundant in the sense that the variables being updated are also updated in the second Metropolis-Hastings update. However, they are cheap because only parts of the likelihood need to be reevaluated.

4.1.4. *Reweighting using importance sampling.* Let $(\alpha^{(j)}, \lambda^{(j)}, \gamma^{(j)})$, $j = 1, \ldots, J$ be a sample from the posterior distribution corresponding to model M_0. We want to use this sample as an importance sample to compute posterior moments for the baseline model M_i. We evaluate, at each posterior draw j, the prior density $f_0(\alpha, \lambda)$ for model M_0 and the prior density $f_i(\alpha, \lambda)$ for the model i for which we want to compute posterior moments. The importance sampling weights are

$$w_{ij} = \frac{f_i(\alpha^{(j)}, \lambda^{(j)})}{f_0(\alpha^{(j)}, \lambda^{(j)})}.$$

Suppose $h(\alpha, \lambda, \gamma, P)$ is a function whose posterior mean we want to compute for model M_i. Assume the posterior mean exists. For example, h could be the indicator function with value 1 whenever the random choice structure P satisfies weak stochastic transitivity and value 0 whenever it does not. In this example, the posterior mean is the posterior probability that P satisfies weak stochastic transitivity, the numerator in the Bayes factor in favour of the model M_i with WST imposed, relative to the baseline model M_i. A simulation consistent approximation of $E[h(\alpha, \lambda, \gamma, P)|M_i]$ is given by

$$\hat{h} \equiv \frac{N}{D} \equiv \frac{\sum_{j=1}^{J} w_{ij} h(\alpha^{(j)}, \lambda^{(j)}, \gamma^{(j)}, P^{(j)})}{\sum_{j=1}^{J} w_{ij}}. \tag{6}$$

We compute an approximation of the variance of $\hat{h}$, random because of simulation noise, in the following way. We first use the batch mean method to approximate variances of the numerator and denominator and their covariance. Then we use the delta method to approximate the variance of the ratio.

The batch mean approximation of the variance σ_N^2 of the numerator is

$$\hat{\sigma}_N^2 \equiv \frac{1}{N_B^2} \sum_{k=1}^{N_B} (\bar{n}_k - \bar{n})^2,$$

where N_B is the number of batches, $B = J/N_B$ is the batch length, (we choose J and N_B so that B is an integer) $\bar{n}_k = B^{-1} \sum_{j=(k-1)B+1}^{kB} w_{ij} h(\alpha^{(j)}, \lambda^{(j)}, \gamma^{(j)}, P^{(j)})$ is the k'th of B numerator batch means and $\bar{n} = \sum_{j=1}^{J} w_{ij} h(\alpha^{(j)}, \lambda^{(j)}, \gamma^{(j)}, P^{(j)})$ is the numerator sample mean.

We obtain similar approximations $\hat{\sigma}_D^2$ and $\hat{\sigma}_{ND}$ of the denominator variance and the covariance. Then the delta method gives the approximation

$$\hat{\sigma}_h^2 = \frac{\hat{\sigma}_N^2 - 2\hat{h}\hat{\sigma}_{ND} + \hat{h}^2 \hat{\sigma}_D^2}{D^2}$$

of the numerical variance of the ratio, $\hat{h}$. We call the square root, $\hat{\sigma}_h$, the numerical standard error of $\hat{h}$.

The Bayes factor in favour of M_i, over M_0, is a posterior mean whose simulation consistent sample counterpart is the denominator D in (6). The variance of its numerical error is approximated by $\hat{\sigma}_D^2$. $\log D$ is a simulation consistent approximation of the variance of the log Bayes factor. The delta method approximation of its variance is $\hat{\sigma}_D^2/D^2$.

We use a similar approach to compute numerical errors for the prior distribution.

5. Results

Here we report results from artificial data simulations testing the correctness of our posterior simulation methods, and do posterior analysis for data from an experiment described in Regenwetter, Dana, and Davis-Stober (2011).

5.1. **Getting it right.** We perform a simulation whose sole purpose is to test the correctness of our posterior simulation methods. This is a purely pre-data exercise, involving only artificial data. The tests described here are similar to those described in Geweke (2004). We draw a sample from the *joint* distribution of hyper-parameters, latent variables *and data*, for an artificial choice experiment where the master set has $n = 3$ elements and all subsets of size two and three are presented exactly once. We complete the specification of the prior by choosing values $a = \tilde{a} = 10$ and $b = \tilde{b} = 0.1$, and complete the specification of the proposal distribution by choosing the value $\pi = 0.5$. We obtain a sample of size $J = 10^6$.

The initial draw is a direct draw from the joint distribution of δ, $\tilde{\delta}$, γ and N, obtained by first drawing hyper-parameters δ and $\tilde{\delta}$ from their prior distribution, then the latent variable vector γ from its conditional distribution given δ and $\tilde{\delta}$, and then data from their (categorical) conditional distribution given γ. Subsequent draws are the output of a Markov chain whose invariant distribution is the joint distribution of δ, $\tilde{\delta}$, γ and N. A single transition of the chain consists of four Metropolis-Hastings updates. Three are the very same updates used to update the posterior distribution. The fourth is a direct draw of N from its conditional distribution given hyper-parameters and latent variables.

If the Markov chain has the correct invariant distribution and if data simulation and posterior simulation are implemented correctly, then a realization of the chain must be a sample of draws from the correct joint distribution, although the draws will be serially dependent. This is a very strong condition that leads to multiple tests of program correctness.

We test 18 hypotheses implied by program correctness. We know that the marginal distributions of δ and $\tilde{\delta}$ are the same as their prior distributions, both $\mathrm{Ga}(10, 0.1)$. At all draws of δ and $\tilde{\delta}$ in the sample, we evaluate indicator functions $1_{[0,q]}(\cdot)$, for nine different values of q. The value of the indicator function is one when its argument is in the interval $[0, q]$ and zero otherwise. The values of q are the quantiles of the $\mathrm{Ga}(10, 0.1)$ distribution corresponding to the nine probabilities $p = 0.1, 0.2, \ldots, 0.8, 0.9$. The nine values of p and q are tabulated in Table 1.

We then compare the sample means of these indicator functions with what their population counterparts should be, namely the probabilities $0.1, 0.2, \ldots, 0.8, 0.9$ themselves. Table 1 shows the results. Column $\hat{p}, \delta$ gives the sample mean of the indicator function $1_{[0,q]}(\delta)$, for each value of q, and the fourth column gives the numerical standard error for $\hat{p}$. Column $\hat{p}, \tilde{\delta}$ gives the sample mean of the indicator function $1_{[0,q]}(\tilde{\delta})$, and the sixth column gives the numerical standard error for $\hat{p}$.

With our large sample size, we obtain very small standard errors. Even so, the sample means are all within a single standard error of the population means, under the null hypothesis that our code works properly. The results fail to reject this hypothesis.

p	q	$\hat{p}, \delta$	NSE	$\hat{p}, \tilde{\delta}$	NSE
0.1	0.6221	0.0999	0.0005	0.1004	0.0005
0.2	0.7289	0.1998	0.0007	0.1998	0.0007
0.3	0.8133	0.2993	0.0008	0.2996	0.0008
0.4	0.8904	0.3995	0.0009	0.3996	0.0009
0.5	0.9669	0.5001	0.0009	0.4992	0.0009
0.6	1.0476	0.6004	0.0009	0.5992	0.0009
0.7	1.1387	0.7000	0.0008	0.6995	0.0009
0.8	1.2519	0.7998	0.0007	0.7997	0.0007
0.9	1.4206	0.8998	0.0005	0.9004	0.0005

TABLE 1. Sample probabilities for “Getting it right” computations

5.2. **Posterior analysis.** In Regenwetter, Dana, and Davis-Stober’s (2011) experiment, 18 undergraduates participated in three different scenarios, denoted here and in that paper by “Cash I”, “Cash II” and “Noncash”. In each scenario, the master set contains $n = 5$ objects, and the objects are lotteries in which a prize is won with a certain probability. In “Cash I”, the probabilities of winning replicate those from a similar experiment by Tversky (1969), designed to elicit intransitive revealed preferences. Prizes are monetary values, adjusted to approximately replicate the purchasing power of the original prizes in Tversky (1969). In “Cash II”, prizes are also monetary. Probabilities and prizes are chosen so that the expected monetary values of the five lotteries were identical. In “Noncash”, the prizes were non-monetary. In each scenario, all 18 participants were presented all ten doubleton subsets of the master set twenty times.

We wish to illustrate the sensitivity of various results to the choice of prior distribution. To this end, we report results for six baseline models, M_1 through M_6, differing only in terms of the prior distribution. Table 2 defines the priors and gives selected moments. The first four columns define the various priors in terms of the hyper-parameters a, $\tilde{a}$, b and $\tilde{b}$ of equation (2). In all cases, the values of b and $\tilde{b}$ are equal. The next three columns give the implied prior mean, variance and standard deviation of the parameter α. The final three columns do the same for the parameter λ. We will denote the prior density for model M_i as $f_i(\alpha, \lambda)$, for $i = 0, 1, \ldots, 6$.

The model M_0 is used for posterior simulation and we do not report results for it. Its prior, also tabulated in Table 2, is chosen for its property that the ratio $f_i(\alpha, \lambda)/f_0(\alpha, \lambda)$ of prior densities is bounded for all $i = 1, \ldots, 6$. This allows us to compute all results using a single posterior sample, for model M_0: results for other priors are computed using importance sampling, as described in Section 4.1.4.

We now discuss inference based on the following simulations. For each of the 18 participants, we generate a posterior sample of size 2.4×10^7, retaining every 300th draw,

	a	$\tilde{a}$	b	$\tilde{b}$	$E[\alpha]$	Var[α]	σ_α	$E[\lambda]$	Var[λ]	σ_λ
M_0	1.0	0.20	1.3500	3.7500	2.1	4.625	2.153	0.747	0.098	0.313
M_1	1.0	0.20	1.2500	1.2500	1.5	1.875	1.369	0.833	0.076	0.275
M_2	1.0	0.60	0.9375	0.9375	1.5	1.406	1.186	0.625	0.144	0.380
M_3	1.0	1.00	0.7500	0.7500	1.5	1.125	1.061	0.500	0.167	0.408
M_4	1.4	0.20	0.9375	0.9375	1.5	1.406	1.186	0.875	0.067	0.259
M_5	1.4	0.60	0.7500	0.7500	1.5	1.125	1.061	0.700	0.140	0.374
M_6	1.8	0.20	0.7500	0.7500	1.5	1.125	1.061	0.900	0.060	0.245

TABLE 2. Prior hyper-parameters and moments of α and λ

for a thinned sample size of 8.0×10^4. We use the model M_0 with prior hyper-parameter values indicated in Table 2. We complete the specification of the proposal distribution by choosing the value $\pi = 0.5$.

Table 3 shows log Bayes factors in favour of the models M_1 through M_6, relative to the model M_0. All of these models are unconstrained, with no axioms imposed. For a given row, differences of log Bayes factors give log Bayes factors in favour of one model over another. Take, for example, the first subject. The log Bayes factor in favour of M_1 over M_2 is $-0.04 - -0.31 = 0.27$, implying a posterior odds ratio of $\exp(0.27) \approx 1.310$. All numerical standard errors in this table are less than 0.015.

For no participant are Bayes factors decisive. Looking across participants, however, some patterns emerge. For most participants, models M_4 and M_6 are the most favoured. These are models where the prior mean of λ is particularly high. For participants 4 and 16, who are exceptional in many ways, they are the least favoured.

Tables 4 and 5 show posterior means μ_i and standard deviations σ_i of the α and λ parameters, respectively. Each row shows results for a different participant in the experiment; the first column gives the participant's identifier. The tables show results for the six models $M_1, \ldots, M_6$. All numerical standard errors in Table 4 and Table 5 are less than 0.25 and 0.03 respectively.

The variation of the posterior mean of α across participants is considerably more pronounced than its variation across models. This is also true of the posterior standard deviation. For some participants, the data are quite informative about α and the posterior mean and standard deviation of α are much smaller than the prior mean and standard deviation. For others, the posterior mean and standard deviation are close to their prior counterparts.

Relative to α, the posterior mean of λ is fairly sensitive to the prior mean. Except for participants 4 and 16, the posterior mean tends to be larger than the prior mean for all models. This is further evidence that most subjects' behaviour is consistent with values of λ much closer to one than to zero. Again, participants 4 and 16 are exceptions.

Tables 6 through 11 give log Bayes factors in favour of restricted models over baseline models. Each of these tables gives results for a single axiom; for example, Table 6 gives log Bayes factors in favour of a restricted model with weak stochastic transitivity imposed

	M_1	M_2	M_3	M_4	M_5	M_6
1	-0.04	-0.31	-0.50	0.05	-0.19	0.11
2	0.14	-0.36	-0.72	0.30	-0.14	0.41
3	0.26	-0.45	-1.06	0.25	-0.40	0.21
4	-0.18	0.60	0.88	-0.22	0.57	-0.28
5	0.24	-0.51	-1.14	0.33	-0.34	0.38
6	0.06	0.28	0.35	0.16	0.38	0.22
7	0.19	-0.55	-1.14	0.37	-0.29	0.49
8	0.24	0.06	-0.27	0.20	0.05	0.14
9	-0.21	-0.63	-0.94	-0.17	-0.55	-0.14
10	0.21	-0.28	-0.69	0.35	-0.07	0.44
11	0.22	0.15	-0.08	0.24	0.19	0.24
12	-0.15	-0.22	-0.29	-0.07	-0.15	-0.01
13	-0.14	-0.54	-0.83	-0.11	-0.45	-0.09
14	0.26	-0.53	-1.19	0.24	-0.48	0.19
15	-0.01	-0.38	-0.66	0.08	-0.25	0.13
16	0.12	0.79	0.94	0.02	0.75	-0.08
17	-0.86	-0.61	-0.56	-0.93	-0.69	-0.96
18	-0.10	-0.60	-0.97	-0.07	-0.50	-0.06

TABLE 3. Log Bayes factors in favour of models M_i, $1, \ldots, 6$, by participant

against a baseline model. Each table reports log Bayes factors for each participant and each baseline model.

Numerical standard errors for log Bayes factors vary greatly. The error tends to be larger for the more improbable axioms and the smallest (i.e. most negative) log Bayes factors, due to the difficulty of measuring the probability of rare events. In the most extreme cases, the log Bayes factor is given as "-Inf", indicating that not a single posterior draw of the random choice structure P, out of 8×10^5, satisfied the relevant axiom. In those cases where there is a great deal of uncertainty about the log Bayes factor, at least we know that it is very small, and that the data strongly favour the baseline model.

For most participants, weak stochastic transitivity is favoured over the baseline model for all six baseline models. The data of participants 4, 6, 12 and 17 stand out as favouring the baseline model in all cases. Whether favourable or unfavourable, log Bayes factors are fairly robust to the prior specification: they do not vary much across models.

Log Bayes factors in favour of WST indicate quite weak support — the largest Bayes factor in favour is $\exp(0.98) \approx 2.66$. This is because the prior probability of WST is so high. For all models, the highest log Bayes factors are very close to each other; for the subjects with the highest Bayes factors, the posterior probabilities of WST are very close to one, giving Bayes factors very close to the upper bound for a particular model, namely the reciprocal of the prior probability of WST under that model. Reported log Bayes factors in favour of WST have numerical standard errors less than 0.05. Excluding participants 4, 12 and 17, they are less than 0.01.

	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4	μ_5	σ_5	μ_6	σ_6
1	2.61	1.35	2.50	1.18	2.41	1.06	2.43	1.19	2.36	1.07	2.30	1.08
2	1.79	0.99	1.77	0.91	1.75	0.84	1.75	0.91	1.73	0.85	1.72	0.85
3	0.51	0.35	0.51	0.33	0.52	0.32	0.57	0.36	0.57	0.34	0.62	0.37
4	1.60	0.76	1.60	0.71	1.60	0.67	1.58	0.73	1.58	0.68	1.56	0.70
5	0.76	0.48	0.78	0.46	0.79	0.44	0.82	0.48	0.83	0.46	0.86	0.47
6	1.55	0.82	1.49	0.72	1.47	0.66	1.55	0.78	1.51	0.70	1.55	0.74
7	1.45	0.81	1.44	0.76	1.44	0.72	1.45	0.77	1.45	0.73	1.46	0.73
8	0.52	0.36	0.49	0.33	0.49	0.31	0.58	0.37	0.56	0.34	0.64	0.38
9	3.03	1.73	2.96	1.53	2.87	1.37	2.69	1.45	2.66	1.32	2.47	1.26
10	1.09	0.64	1.09	0.60	1.09	0.57	1.13	0.62	1.13	0.58	1.16	0.60
11	0.67	0.43	0.63	0.39	0.62	0.37	0.73	0.43	0.69	0.40	0.79	0.44
12	2.45	1.31	2.47	1.19	2.44	1.08	2.27	1.16	2.30	1.07	2.15	1.05
13	3.17	1.67	2.98	1.44	2.84	1.28	2.85	1.42	2.73	1.27	2.63	1.25
14	0.49	0.34	0.50	0.33	0.51	0.32	0.55	0.35	0.56	0.34	0.60	0.36
15	2.77	1.39	2.59	1.20	2.46	1.07	2.57	1.22	2.44	1.09	2.42	1.10
16	0.76	0.44	0.78	0.42	0.79	0.40	0.83	0.45	0.84	0.43	0.88	0.45
17	3.05	1.75	3.07	1.53	2.96	1.35	2.63	1.44	2.70	1.32	2.38	1.25
18	3.31	1.69	3.05	1.45	2.87	1.28	2.96	1.44	2.80	1.27	2.73	1.27

TABLE 4. Posterior mean μ_i and standard deviation σ_i of α, by participant and model M_i, $i = 1, \ldots, 6$

Log Bayes factors in favour of moderate stochastic transitivity vary considerably. The empirical evidence against MST is strong and robust for participants 4, 16 and 17. Where the Bayes factors favour MST, the degree of support is often stronger than for WST. This is possible because of the relative prior improbability of MST. In some cases, of course, the evidence turns the other way. The data for participant 16 is quite consistent with WST but not with MST, a stronger condition. Note that weak evidence against WST for participant 6 becomes weak evidence in favour of MST. Estimated log Bayes factors in favour of MST have numerical standard errors less than 0.5. Excluding participants 4, 16 and 17, they are less than 0.05.

For participants 4, 16 and 17, not a single posterior draw satisfies strong stochastic transitivity. This makes it impossible to measure the log Bayes factor with any accuracy, but we do know that it is very small. There is clearly strong evidence against SST for these participants. For other participants, there is moderate and robust evidence in favour of SST. Excluding participants 4, 16 and 17, estimated log Bayes factor in favour of SST have numerical standard errors less than 1.1. Excluding 9 and 12 as well, they are less than 0.5.

As with WST, log Bayes factors in favour of the triangle inequality are small. Only the log Bayes factors for model M_3 exceed one, and we have seen that this model has relatively weak empirical support for participants other than 4 and 16. Again, weak support is due to the relatively high prior probability of the axiom in question. Estimated log Bayes factor

	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4	μ_5	σ_5	μ_6	σ_6
1	0.91	0.17	0.75	0.25	0.63	0.27	0.93	0.14	0.80	0.21	0.94	0.12
2	0.95	0.11	0.85	0.19	0.75	0.22	0.96	0.10	0.87	0.16	0.96	0.09
3	0.95	0.11	0.86	0.18	0.78	0.21	0.96	0.09	0.88	0.15	0.97	0.08
4	0.59	0.28	0.49	0.27	0.42	0.25	0.65	0.25	0.56	0.25	0.70	0.23
5	0.97	0.08	0.90	0.14	0.82	0.18	0.97	0.07	0.91	0.12	0.97	0.06
6	0.82	0.24	0.65	0.27	0.54	0.27	0.85	0.20	0.70	0.24	0.88	0.18
7	0.97	0.07	0.91	0.12	0.84	0.16	0.97	0.06	0.92	0.11	0.98	0.06
8	0.87	0.18	0.76	0.22	0.68	0.23	0.90	0.15	0.80	0.19	0.92	0.13
9	0.92	0.17	0.77	0.25	0.66	0.27	0.94	0.13	0.82	0.20	0.95	0.11
10	0.95	0.10	0.86	0.17	0.78	0.20	0.96	0.09	0.87	0.15	0.96	0.08
11	0.87	0.19	0.74	0.23	0.66	0.24	0.89	0.16	0.78	0.20	0.91	0.14
12	0.87	0.22	0.68	0.28	0.56	0.28	0.90	0.18	0.75	0.25	0.92	0.15
13	0.92	0.16	0.78	0.24	0.66	0.26	0.93	0.13	0.82	0.20	0.94	0.11
14	0.96	0.10	0.87	0.17	0.79	0.20	0.96	0.08	0.89	0.14	0.97	0.07
15	0.93	0.15	0.79	0.22	0.68	0.25	0.94	0.13	0.82	0.19	0.95	0.11
16	0.62	0.25	0.55	0.25	0.49	0.24	0.67	0.22	0.60	0.22	0.71	0.20
17	0.73	0.30	0.56	0.30	0.47	0.28	0.81	0.24	0.65	0.27	0.85	0.20
18	0.93	0.14	0.81	0.22	0.70	0.25	0.94	0.12	0.84	0.19	0.95	0.10

TABLE 5. Posterior mean μ_i and standard deviation σ_i of λ, by participant and model M_i, $i = 1, \ldots, 6$

in favour of TI have numerical standard errors less than 0.06. Excluding partcipants 4 and 16, they are less than 0.02.

Results for regularity are similar to those for TI. Estimated log Bayes factor in favour of Reg have numerical standard errors less than 0.05. Excluding partcipants 4 and 16, they are less than 0.02.

There is more support for the random ranking hypothesis than there is for TI. This is possible because of its lower prior probability. Overall, there is considerable support for RR, more so than that for TI, a necessary condition for RR. This evidence is fairly robust across baseline models, so that even when the prior distribution puts more mass on values of λ close to one, there is still almost as much of an improvement in out-of-sample predictive performance resulting from imposing RR. Even for participants 4 and 16, the evidence against RR is weaker than the evidence against TI. It seems that the conditional posterior probability of RR given TI is considerably higher than the corresponding conditional prior probability.

A remarkable feature of the results in this table is that for participants 4 and 16, the Bayes factor *favours* RR for the baseline model M_3. Recall that this model is the best performing baseline model for these two subjects and the worst for most of the other participants. It is the model for which the prior mean of λ, at 0.5, is the lowest, giving a particularly low prior probability of RR. However, truncating to the RR region improves the predictive performance.

	M_1	M_2	M_3	M_4	M_5	M_6
1	0.02	0.27	0.43	-0.02	0.18	-0.05
2	0.38	0.73	0.97	0.33	0.63	0.30
3	0.38	0.74	0.98	0.33	0.63	0.30
4	-3.44	-3.93	-4.00	-3.30	-3.88	-3.19
5	0.38	0.74	0.98	0.33	0.63	0.30
6	-0.39	-0.33	-0.24	-0.40	-0.39	-0.40
7	0.38	0.74	0.98	0.33	0.63	0.30
8	0.38	0.74	0.98	0.33	0.63	0.30
9	0.03	0.25	0.39	-0.00	0.19	-0.02
10	0.38	0.74	0.98	0.33	0.63	0.30
11	0.38	0.74	0.98	0.33	0.63	0.30
12	-0.81	-0.79	-0.78	-0.82	-0.82	-0.83
13	-0.10	0.13	0.28	-0.15	0.05	-0.17
14	0.38	0.74	0.98	0.33	0.63	0.30
15	0.29	0.62	0.84	0.24	0.52	0.21
16	0.34	0.69	0.93	0.29	0.59	0.26
17	-1.91	-2.26	-2.40	-1.82	-2.17	-1.76
18	0.10	0.42	0.62	0.06	0.32	0.03

TABLE 6. Log Bayes factors in favour of weak stochastic transitivity, by participant and model

Results for the multiplicative inequality are somewhat unusual. For participants 3,4,5,10,11 and 16, there is strong evidence against MI for all baseline models. For other participants, prior and posterior probabilities are both very low, and the log Bayes factors are measured with a lot of error: numerical standard errors for them range from 0.3 to 1.1. Even considering the large standard errors, however, there seems to be a lot of sensitivity to the prior specification.

6. Conclusions

We have introduced new posterior simulation methods allowing more flexible inference for random choice structures. Previous articles had specified prior distributions over the set of binary choice probabilities in which the probabilities were independent, each with a Beta distribution. Such priors are a convenient choice, since they are fully conjugate for the likelihood function for choices that are independent across choice sets and trials. However, ruling out prior dependence is quite restrictive. Our methods work for the two-parameter class of prior distributions introduced in McCausland and Marley (2013). The α parameter governs consistency of choice from trial to trial and λ governs dependence of choice probabilities across choice sets.

	M_1	M_2	M_3	M_4	M_5	M_6
1	0.41	0.76	1.01	0.35	0.64	0.31
2	1.03	1.55	1.91	0.97	1.40	0.93
3	0.85	1.38	1.75	0.80	1.24	0.76
4	-6.52	-7.74	-7.58	-6.37	-7.85	-6.25
5	0.87	1.42	1.79	0.83	1.29	0.80
6	0.10	0.18	0.31	0.09	0.12	0.09
7	0.30	0.81	1.15	0.25	0.66	0.21
8	1.08	1.54	1.86	1.05	1.42	1.02
9	-1.43	-1.27	-1.20	-1.51	-1.40	-1.57
10	1.12	1.72	2.13	1.07	1.57	1.04
11	1.35	1.93	2.31	1.29	1.78	1.25
12	-1.97	-1.84	-1.79	-2.03	-1.92	-2.08
13	0.59	0.95	1.21	0.52	0.82	0.46
14	0.98	1.50	1.85	0.92	1.36	0.89
15	1.51	2.02	2.37	1.45	1.87	1.41
16	-4.70	-5.61	-6.07	-4.62	-5.55	-4.55
17	-4.47	-5.03	-5.26	-4.42	-4.97	-4.41
18	1.12	1.60	1.92	1.05	1.46	1.00

TABLE 7. Log Bayes factors in favour of moderate stochastic transitivity, by participant and model

We have shown that for most participants in an experiment, there is strong evidence for dependence across choices. Data are quite informative about the degree of dependence, as measured by λ.

Certain broad inferences are fairly robust to the prior specification. For all but two participants, there is weak evidence for weak stochastic transitivity and the triangle inequality. Overall, there is more support for the random ranking hypothesis than there is for the triangle inequality, a necessary condition. Evidence for and against other axioms varies more by participant. Evidence against the multiplicative inequality is quite strong for a large fraction of the participants.

7. Acknowledgements

The authors gratefully acknowledge funding by the Social Sciences and Humanities Research Council of Canada in the form of Insight grant SSHRC 435-2012-0451 to the University of Victoria for Marley and McCausland.

They also thank Clintin Davis-Stober for providing experimental data.

Appendix A. Markov chain Monte Carlo details

A.1. **Transition densities.** We define some proposal distributions we use in our Metropolis-Hastings updates. The first is the Beta-Gamma transformation introduced in Lewis,

	M_1	M_2	M_3	M_4	M_5	M_6
1	0.47	1.51	1.90	1.14	1.52	1.22
2	-1.53	0.00	0.68	-0.82	-0.00	-0.70
3	-0.13	1.08	1.55	0.62	1.10	0.76
4	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
5	-0.49	0.78	1.29	0.28	0.85	0.44
6	1.01	2.01	2.45	1.76	2.07	1.90
7	-0.85	0.70	1.51	-0.14	0.67	-0.03
8	0.07	1.26	1.86	0.78	1.23	0.90
9	-3.52	-2.98	-3.66	-2.77	-2.85	-2.66
10	-0.27	1.10	1.78	0.48	1.13	0.63
11	1.08	2.56	3.29	1.75	2.49	1.83
12	-4.18	-7.46	-12.08	-3.29	-7.19	-3.05
13	0.10	1.00	1.34	0.75	0.98	0.82
14	-0.19	0.93	1.48	0.53	0.90	0.65
15	2.31	3.59	4.24	3.00	3.57	3.09
16	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
17	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
18	1.54	2.63	3.10	2.23	2.62	2.33

TABLE 8. Log Bayes factors in favour of strong stochastic transitivity, by participant and model

McKenzie, and Hugus (1986). Suppose we transform a random variable x to create $x^* = \beta x + \epsilon$, where x, β and ϵ are mutually independent, $\beta \sim \mathrm{Be}(\pi a, (1-\pi))$, $\epsilon \sim \mathrm{Ga}((1-\pi)a)$. Here, $\pi \in (0,1)$ and $a > 0$ are parameters. The unique invariant distribution of this transformation is $\alpha \sim \mathrm{Ga}(a)$. We denote the transition density as $q_1(x^*|x, a, \pi)$. The Markov chain with this transition density is known as the Beta-Gamma autoregressive process. Importantly, Lewis, McKenzie, and Hugus (1986) show that it is time reversible, which implies that

$$f_{\mathrm{Ga}}(\alpha|a)q_1(\alpha^*|\alpha, a, \pi) = f_{\mathrm{Ga}}(\alpha^*|a)q_1(\alpha|\alpha^*, a, \pi). \tag{7}$$

We will never need to evaluate $q_1(\cdot, \cdot)$ but we will need to invoke the time reversibility condition (7) to demonstrate the correctness of our Metropolis-Hastings updates.

We now derive the transition density $q_2(y^*|y, x, x^*)$ for a conditional transformation from y to y^* given x and x^*, where $x > 0$ and $x^* > 0$, $x \neq x^*$, are parameters. The transformation is defined as follows. If $x^* > x$, then $y^* = y + \epsilon$, where ϵ and (y, x, x^*) are independent and $\epsilon \sim \mathrm{Ga}(x^* - x)$. If $x^* < x$, then $y^* = \beta y$, where $\beta \sim \mathrm{Be}(x^*, x - x^*)$.

The conditional density associated with the conditional transition from y to y^* given x and x^* is

$$q_2(y^*|y, x, x^*) = \begin{cases} f_{\mathrm{Ga}}(y^* - y|x^* - x) & x^* > x, \\ \frac{1}{y} f_{\mathrm{Be}}\left(\frac{y^*}{y}|x^*, x - x^*\right) & x > x^*, \end{cases}$$

	M_1	M_2	M_3	M_4	M_5	M_6
1	0.38	0.93	1.30	0.32	0.80	0.28
2	0.40	0.97	1.33	0.33	0.83	0.29
3	0.28	0.65	0.87	0.23	0.55	0.19
4	-2.21	-2.74	-2.78	-2.08	-2.69	-1.97
5	0.34	0.81	1.09	0.28	0.68	0.24
6	0.03	0.29	0.54	0.00	0.20	-0.01
7	0.40	0.95	1.29	0.33	0.80	0.28
8	0.10	0.41	0.64	0.08	0.33	0.07
9	0.38	0.93	1.28	0.32	0.80	0.28
10	0.36	0.88	1.22	0.30	0.74	0.26
11	0.19	0.62	0.93	0.16	0.51	0.14
12	0.21	0.52	0.73	0.17	0.43	0.15
13	0.40	0.98	1.36	0.34	0.85	0.29
14	0.30	0.69	0.91	0.24	0.58	0.21
15	0.44	1.07	1.49	0.37	0.92	0.32
16	-3.19	-4.33	-5.03	-3.09	-4.25	-3.00
17	-0.17	-0.07	0.06	-0.14	-0.10	-0.13
18	0.44	1.08	1.51	0.37	0.93	0.33

TABLE 9. Log Bayes factors in favour of triangle inequality, by participant and model

where f_{Ga} denotes the standard Gamma density,

$$f_{\mathrm{Ga}}(y|x) = \frac{y^{x-1}}{\Gamma(x)}, \quad x > 0,\ y > 0,$$

and f_{Be} denotes the Beta density,

$$f_{\mathrm{Be}}(y|x_1, x_2) = \frac{\Gamma(x_1 + x_2)}{\Gamma(x_1)\Gamma(x_2)} y^{x_1-1}(1-y)^{x_2-1}, \quad x_1, x_2 > 0,\ y \in (0, 1).$$

We now show an important result:

$$f_{\mathrm{Ga}}(y|x)q(y^*|y, x, x^*) = f_{\mathrm{Ga}}(y^*|x^*)q(y|y^*, x^*, x). \tag{8}$$

Proof: Write out the left hand side of (8) as

$$\begin{aligned} f_{\mathrm{Ga}}(y|x)&q(y^*|y, x, x^*) \\ &= \frac{y^{x-1}}{\Gamma(x)}\left[u(x^*-x)\frac{(y^*-y)^{x^*-x-1}}{\Gamma(x^*-x)} + u(x-x^*)\frac{1}{y}\frac{\Gamma(x)}{\Gamma(x^*)\Gamma(x-x^*)}\left(\frac{y^*}{y}\right)^{x^*-1}\left(\frac{y-y^*}{y}\right)^{x-x^*-1}\right] \\ &= u(x^*-x)\frac{y^{x-1}(y^*-y)^{x^*-x-1}}{\Gamma(x)\Gamma(x^*-x)} + u(x-x^*)\frac{(y^*)^{x^*-1}(y-y^*)^{x-x^*-1}}{\Gamma(x^*)\Gamma(x-x^*)}, \end{aligned}$$

where $u(\cdot)$ is the Heaviside, or unit step function, equal to one for non-negative arguments and zero for negative arguments.

	M_1	M_2	M_3	M_4	M_5	M_6
1	0.20	0.48	0.68	0.16	0.40	0.14
2	0.34	0.87	1.27	0.28	0.74	0.25
3	0.33	0.87	1.34	0.28	0.77	0.26
4	-2.28	-3.06	-3.33	-2.11	-2.94	-1.98
5	0.41	1.09	1.64	0.35	0.94	0.31
6	-0.28	-0.45	-0.49	-0.26	-0.46	-0.25
7	0.45	1.22	1.85	0.39	1.05	0.34
8	-0.18	-0.13	0.10	-0.16	-0.14	-0.15
9	0.25	0.65	0.96	0.23	0.59	0.21
10	0.29	0.76	1.16	0.23	0.63	0.20
11	-0.19	-0.19	-0.02	-0.17	-0.20	-0.16
12	0.06	0.15	0.24	0.06	0.14	0.05
13	0.23	0.60	0.89	0.20	0.52	0.18
14	0.36	0.97	1.47	0.32	0.85	0.28
15	0.25	0.63	0.92	0.20	0.53	0.17
16	-3.07	-3.86	-4.08	-2.97	-3.82	-2.88
17	-0.46	-0.72	-0.77	-0.35	-0.59	-0.28
18	0.30	0.76	1.14	0.25	0.66	0.22

TABLE 10. Log Bayes factors in favour of regularity, by participant and model

The last line has the symmetry property that replacing (x, y) by (x^*, y^*) gives the same expression. The left hand side must have the same property, which is the desired result.

A.2. **Hastings ratios.** Writing out the full Hastings ratio for the first Metropolis-Hastings update gives

$$H = \frac{f(\alpha^*) \prod_{\succ \in R(T)} f_{\mathrm{Ga}}(\gamma^*(\succ)|\alpha^*/n!) \Pr[N|\gamma^*]}{f(\alpha) \prod_{\succ \in R(T)} f_{\mathrm{Ga}}(\gamma(\succ)|\alpha/n!) \Pr[N|\gamma]} \cdot \frac{q_1(\alpha|\alpha^*) \prod_{\succ \in R(T)} q_2(\gamma(\succ)|\gamma^*(\succ), \alpha^*/n!, \alpha/n!)}{q_1(\alpha^*|\alpha) \prod_{\succ \in R(T)} q_2(\gamma^*(\succ)|\gamma(\succ), \alpha/n!, \alpha^*/n!)},$$

where γ^* is understood to mean the vector γ of all weights, with the $\gamma(\succ)$ weights replaced by $\gamma^*(\succ)$, $\succ \in R$.

Using equation (7) and repeated applications of equation (8), the Hastings ratio reduces to

$$H = \frac{\Pr[N|\gamma^*]}{\Pr[N|\gamma]}.$$

Therefore the first Metropolis-Hastings update is a true Metropolis-Hastings update preserving the conditional distribution of δ and $\gamma(\succ)$, $\succ \in R(T)$, given $\tilde{\delta}$, other latent variables, and data N. The analogous demonstration for the second Metropolis-Hastings update is very similar and we omit it.

	M_1	M_2	M_3	M_4	M_5	M_6
1	1.35	1.60	1.91	1.29	1.49	1.25
2	1.51	2.02	2.46	1.42	1.85	1.37
3	1.50	2.05	2.56	1.43	1.91	1.39
4	-1.08	-1.05	1.34	-0.97	-1.75	-0.86
5	1.58	2.25	2.80	1.49	2.06	1.43
6	0.89	0.73	1.21	0.88	0.65	0.87
7	1.62	2.36	2.96	1.53	2.16	1.47
8	1.00	1.09	1.86	0.98	0.99	0.98
9	1.40	1.76	2.09	1.35	1.66	1.31
10	1.46	1.93	2.39	1.37	1.75	1.32
11	0.98	1.02	1.72	0.97	0.93	0.97
12	1.22	1.30	1.57	1.19	1.26	1.17
13	1.38	1.70	1.95	1.32	1.59	1.28
14	1.54	2.15	2.81	1.46	1.98	1.41
15	1.40	1.73	1.96	1.33	1.60	1.28
16	-1.82	-1.13	1.50	-1.82	-2.42	-1.75
17	0.69	0.42	0.92	0.78	0.48	0.82
18	1.44	1.84	2.16	1.37	1.71	1.32

TABLE 11. Log Bayes factors in favour of random ranking, by participant and model

REFERENCES

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, Second Edition*. Springer-Verlag, New York, NY.

Bernardo, J. M., and Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley and Sons, Chichester, England.

Block, H. D., and Marschak, J. (1960). 'Random Orderings and Stochastic Theories of Responses', in I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 97–132. Stanford University Press, Stanford, CA.

Busemeyer, J. R., and Rieskamp, J. (2013). 'Psychological research and theories on preferential choice', in S. Hess and A. Daly (eds.), *Handbook of Choice Modelling*. Edward Elgar.

Cavagnaro, D. R., and Davis-Stober, C. P. (2013). 'Transitive in our Preferences, but Transitive in Different Ways: An Analysis of Choice Variability', Manuscript.

Chib, S., and Greenberg, E. (1995). 'Understanding the Metropolis Hastings Algorithm', *American Statistical Journal*, 49: 327–335.

Colonius, H. (1983). 'A characterization of stochastic independence by association, with an application to random utility theory', *Journal of Mathematical Psychology*, 27(1): 103–105.

Dagsvik, J. K. (1994). 'Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes', *Econometrica*, 62(5): 1179–1205.

Dridi, T. (1980). 'Sur les distributions binaires associées à des distributions ordinales.', *Mathématiques et Sciences Humaines*, 69: 15–31.

Fiorini, S. (2004). 'A short proof of a theorem of Falmagne', *Journal of Mathematical Psychology*, 48: 80–82.

Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2011). *Statistical Distributions*. John Wiley & Sons.

Geweke, J. (1989). 'Bayesian Inference in Econometric Models Using Monte Carlo Integration', *Econometrica*, 57: 1317–1340.

	M_1	M_2	M_3	M_4	M_5	M_6
1	-1.69	-0.70	0.54	-0.05	1.32	1.52
2	-3.65	-2.18	-1.56	-1.69	-0.03	0.11
3	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
4	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
5	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
6	-2.19	-1.27	-0.37	-0.20	1.02	1.62
7	-3.72	-4.86	-6.85	-1.81	-2.78	-0.06
8	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
9	-1.28	-0.70	-0.17	0.01	0.87	1.28
10	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
11	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
12	-1.74	-0.36	0.99	-0.06	1.78	1.56
13	-1.64	-1.03	-0.94	-0.11	0.72	1.37
14	-2.50	-1.73	-1.43	-1.65	-0.59	-0.93
15	-2.12	-0.33	0.86	-0.08	1.92	1.77
16	-Inf	-Inf	-Inf	-Inf	-Inf	-Inf
17	-1.59	-1.24	-0.81	0.50	1.12	2.35
18	-0.90	0.33	1.30	0.33	1.91	1.52

TABLE 12. Log Bayes factors in favour of multiplicative inequality, by participant and model

Geweke, J. (2004). 'Getting it Right: Joint Distribution Tests of Posterior Simulators', *Journal of the American Statistical Association*, 99: 799–804.

——— (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley, Hoboken, New Jersey.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, Boca Raton.

Iverson, G. J., and Falmagne, J.-C. (1985). 'Statistical Issues in Measurement', *Mathematical Social Sicences*, 10: 131–153.

Koppen, M. (1995). 'Random utility representations of binary choice probabilities: Critical graphs yielding critical necessary conditions', *Journal of Mathematical Psychology*, 39: 21–39.

Lewis, P., McKenzie, E., and Hugus, D.K., D. o. O. R. (1986). *Gamma Processes*. Naval Postgraduate School, Monterey, CA.

Luce, R. D. (1977). 'The Choice Axiom after Twenty Years', *Journal of Mathematical Psychology*, 15: 215–233.

Luce, R. D. (1994). 'Thurstone and sensory scaling: then and now', *Psychological Review*, 107: 271–277.

Luce, R. D., and Suppes, P. (1965). 'Preference, Utility, and Subjective Probability', in R. D. Luce, R. R. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 3, chap. 19, pp. 249–410. John Wiley & Sons, New York, NY.

Marley, A. A. J. (1992a). 'A selective review of recent characterizations of stochastic choice models using distribution and functional equation techniques.', *Mathematical Social Sicences*, 23: 5–29.

——— (1992b). 'Stochastic models of choice and reaction time: New developments', *Mathematical Social Sicences*, 23: 1–3, 147–149, 251–253.

——— (2002). 'Random utility models and their applications: recent developments', *Mathematical Social Sicences*, 43: 289–302.

McCausland, W. J., and Marley, A. A. J. (2013). 'Prior Distributions for Random Choice Structures', *Journal of Mathematical Psychology*, forthcoming.
McFadden, D. (1977). 'Modelling the Choice of Residential Location', Cowles Foundation Discussion Papers 477, Cowles Foundation for Research in Economics, Yale University.
McFadden, D., and Train, K. (2000). 'Mixed MNL models for discrete response', *Journal of Applied Econometrics*, 15: 447–470.
Meyn, S. P., and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London.
Myung, J. I., Karabatsos, G., and Iverson, G. J. (2005). 'A Bayesian approach to testing decision making axioms', *Journal of Mathematical Psychology*.
Regenwetter, M., Dana, J., and Davis-Stober, C. P. (2011). 'Transitivity of Preferences', *Psychological Review*, 118: 42–56.
Rieskamp, J., Busemeyer, J. R., and Mellers, B. A. (2006). 'Extending the bounds of rationality: Evidence and theories of preferential choice', *Journal of Economic Literature*, 44: 631–661.
Robert, C. P., and Casella, G. (2010). *Monte Carlo Statistical Methods*. Springer.
Sattath, S., and Tversky, A. (1976). 'Unite and Conquer: A Multiplicative Inequality for Choice Probabilities', *Econometrica*, 44: 79–89.
Suck, R. (2002). 'Independent random utility representations', *Mathematical Social Sicences*, 43: 371–389.
Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edn.
Tversky, A. (1969). 'Intransitivity of Preferences', *Psychological Review*, 76(1): 31–48.
——— (1972). 'Elimination by Aspects: A Theory of Choice', *Psychological Review*, 79(4): 281–299.
Zwilling, C., Cavagnaro, D., and 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.

Département de sciences économiques and CIREQ, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7, Canada, and CIRANO and CenSoC.
E-mail address: william.j.mccausland@umontreal.ca

Department of Psychology, University of Victoria, P.O Box 3050 Victoria BC V8W 3P5, Canada, and CenSoC.
E-mail address: ajmarley@uvic.ca

Récents cahiers de recherche du CIREQ
Recent Working Papers of CIREQ

Si vous désirez obtenir des exemplaires des cahiers, vous pouvez les télécharger à partir de notre site Web http://www.cireqmontreal.com/cahiers-de-recherche

If you wish to obtain copies of the working papers, you can download them directly from our website, http://www.cireqmontreal.com/cahiers-de-recherche

09-2012 Bossert, W., K. Suzumura, "Multi-Profile Intertemporal Social Choice", juillet 2012, 20 pages

10-2012 Bakis, O., B. Kaymak, "On the Optimality of Progressive Income Redistribution", août 2012, 42 pages

11-2012 Poschke, M., "The Labor Market, the Decision to Become an Entrepreneur, and the Firm Size Distribution", août 2012, 29 pages

12-2012 Bossert, W., Y. Sprumont, "Strategy-proof Preference Aggregation", août 2012, 23 pages

13-2012 Poschke, W., "Who Becomes an Entrepreneur? Labor Market Prospects and Occupational Choice", septembre 2012, 49 pages

14-2012 Benchekroun, H., G. Gaudet, H. Lohoues, "Some Effects of Asymmetries in a Common Pool Natural Resource Oligopoly", août 2012, 24 pages

15-2012 Ehlers, L., B. Klaus, "Strategy-Proofness Makes the Difference : Deferred-Acceptance with Responsive Priorities", septembre 2012, 32 pages

16-2012 Bossert, W., C.X. Qi, J.A. Weymark, "An Axiomatic Characterization of the MVSHN Group Fitness Ordering", septembre 2012, 20 pages

17-2012 Ruge-Murcia, F., "Skewness Risk and Bond Prices", mai 2012, 41 pages

18-2012 Amarante, M., M. Ghossoub, E. Phelps, "Contracting for Innovation under Knightian Uncertainty", septembre 2012, 37 pages

19-2012 Bossert, W., C. D'Ambrosio, "Proximity-Sensitive Individual Deprivation Measures", décembre 2012, 15 pages

01-2013 Bossert, W., Y. Sprumont, "Every Choice Function is Backwards-Induction Rationalizable", janvier 2013, 15 pages

02-2013 Amarante, M., "Conditional Expected Utility", février 2013, 19 pages

03-2013 Benchekroun, H., F. Taherkhani, "Adaptation and the Allocation of Pollution Reduction Costs", mai 2013, 33 pages

04-2013 Bossert, W., H. Peters, "Single-Basined Choice", juin 2013, 15 pages

05-2013 Okasha, S., J.A. Weymark, W. Bossert, "Inclusive Fitness Maximization : An Axiomatic Approach", mai 2013, 17 pages

06-2013 Ehlers, L., B. Klaus, "House Allocation via Deferred-Acceptance", juillet 2013, 13 pages