Available online at www.sciencedirect.com
SCIENCE DIRECT

Journal of Mathematical Psychology 49 (2005) 51–69

Journal of Mathematical Psychology

www.elsevier.com/locate/jmp

The exchangeable multinomial model as an approach to testing deterministic axioms of choice and measurement

George Karabatsos*

University of Illinois-Chicago, 1040 W. Harrison St. (M/C 147), Chicago, IL 60607-7133, USA

Received 16 April 2003; received in revised form 20 February 2004
Available obline 16 December 2004

Abstract

The multinomial (Dirichlet) model, derived from de Finetti's concept of exchangeability, is proposed as a general Bayesian framework to test axioms on data, in particular, deterministic axioms characterizing theories of choice or measurement. For testing, the proposed framework does not require a deterministic axiom to be cast in a probabilistic form (e.g., casting deterministic transitivity as weak stochastic transitivity). The generality of this framework is demonstrated through empirical tests of 16 different axioms, including transitivity, consequence monotonicity, segregation, additivity of joint receipt, stochastic dominance, coalescing, restricted branch independence, double cancellation, triple cancellation, and the Thomsen condition. The model generalizes many previously proposed methods of axiom testing under measurement error, is analytically tractable, and provides a Bayesian framework for the random relation approach to probabilistic measurement (J. Math. Psychol. 40 (1996) 219). A hierarchical and nonparametric generalization of the model is discussed.

Keywords: Axiom testing; Exchangeability; Bayesian analysis

1. Introduction

Many models of choice and measurement can be characterized by a set of qualitative axioms. A set of axioms is a set of assumptions that are jointly sufficient for a model's representation, and also, jointly sufficient for a given level of the uniqueness of the representation. At the same time, any single axiom clarifies what predictions a model makes about subjects' behavior.

In many cases, an axiom states a set of binary order relations that must hold over elements of a non-empty set $A = \{a, b, c, \ldots\}$ (where A is possibly a Cartesian-product). The elements refer to, for example, different objects, stimuli, gambles, experimental conditions, or groups of subjects. For any single pair of elements $\{a, b\}$, a binary relation is of the form $a \succsim b, a \sim b$, or $a \succ b$, where $\succ$ denotes "preference" (or "dominance"), $\sim$ denotes "indifference", and $\succsim$ refers to "preference or indifference". Many axioms have been developed for theories of choice and measurement, and virtually all of the empirically testable axioms can be classified as belonging to one of four types. Table 1 contains a list of some of the names of well-known deterministic axioms that correspond to each type. Table 1 also includes names of some probabilistic axioms.

Given the fundamental scientific implications of axioms, it is of primary importance to test whether data conform to their predictions. However, there is a challenging incompatibility between empirical data and the nature of the axioms. On the one hand, empirical data in general contain random error, which is attributable to a number of sources, such as the inherent unreliability of human (or animal) behavior, and sampling error. On the other hand, the axioms do not account for random error, because they are stated in a deterministic, qualitative form (aside from the probabilistic axioms listed in Table 1). Therefore, there have been major efforts to develop statistical methods of axiom testing, starting with Iverson and Falmagne

*Fax: +1 312 996 5651.
E-mail address: georgek@uic.edu.

0022-2496/$ - see front matter
doi:10.1016/j.jmp.2004.11.001

Table 1
Four different types of axioms, and well-known examples of them

Axiom type	Examples of testable axioms
$a \sim b$	Coalescing; segregation; gain-loss decomposition; additivity of joint receipt; duplex decomposition; commutativity; complementarity.
$a \succsim b \Leftrightarrow c \succsim d$	Order independence; consequence monotonicity; joint independence; order independence of events; monotonicity of joint receipt; stochastic dominance. *Strict form* (replacing $\succsim$ with $\succ$): restricted branch independence; distribution independence; monotonicity of event inclusion. *Probabilistic*: quadruple condition.
$a \succsim b$ and...and $c \succsim d \Rightarrow e \succsim f$	Transitivity; double, triple, quadruple, and higher-order cancellation; distributive, and dual-distributive cancellation; stochastic dominance. *Strict form* (replacing $\succsim$ with $\succ$): Lower cumulative independence; upper cumulative independence. *Probabilistic*: Bi-cancellation, weak stochastic transitivity, moderate stochastic transitivity, strong stochastic transitivity.
$a \sim b$ and...and $c \sim d \Rightarrow e \sim f$	Thomsen condition, gains partition, *N*-compound invariance.

(1985) likelihood-ratio method, and continuing with the Bayesian statistical methods developed by Karabatsos (Karabatsos, 2001, 2005; Karabatsos & Ullrich, 2003; Karabatsos & Sheu, 2004).

The present study continues these efforts by proposing a Bayesian statistical framework, described in Section 2. This framework provides a practical basis for testing any deterministic axiom, on discrete- or real-valued response data that can be represented as frequencies in one or more multidimensional contingency tables. In particular, the data are modeled with a Dirichlet posterior distribution, under a multinomial sampling distribution for the data, and a Dirichlet prior distribution specified over the multinomial parameters. This *exchangeable multinomial model* is motivated from de Finetti's (1930, 1937/1964, 1970/1974, 1970/1975) representation theorem, based on his concept of exchangeability for 0–1 random vectors, each random vector representing a multinomial response (see Bernardo (1996) for a recent discussion of this concept). Furthermore, within this framework, a Bayes factor can be directly computed to provide tests of any type of deterministic axiom. Section 3 applies the Bayesian framework to tests of 16 different axioms on real data. All four types of axioms defined in Table 1 are exemplified in these tests. The conclusions of Section 4 explain how this exchangeable multinomial model generalizes models that have been previously proposed for axiom testing. Moreover, this more general model is far easier to implement, and it does not require a deterministic axiom to be cast in a probabilistic form (e.g., casting deterministic transitivity as weak stochastic transitivity). A hierarchical and nonparametric generalization of the exchangeable multinomial model is also discussed.

2. Exchangeable multinomial model: mathematical background

The following four subsections explain the proposed Bayesian framework for axiom testing. In the first subsection, it is shown how each observed response (data point) from a given experiment can be characterized as a 0–1 random vector (i.e., multinomial response). In the second subsection, a multinomial (Dirichlet) model is derived through de Finetti's representation theorem for an exchangeable sequence of 0–1 random vectors, and the third subsection presents a Bayes factor in closed form, which is useful for testing deterministic axioms. Finally, the last subsection presents how the exchangeable model and the Bayes factor are extended to multiple, partially exchangeable 0–1 random vectors.

2.1. Data represented as response patterns

Suppose that in a given experiment, each subject is viewed as having J possible response patterns. These possible response patterns are represented by a J-length vector $R = (r_1, \ldots, r_j, \ldots, r_J)$, where $r_j \neq r_h$ for all pairs $j,\ h \in \{1, \ldots, J\}$, with $j \neq h$. In particular, each r_j denotes a particular response pattern of one or more binary preferences, each binary preference represented as a member of the set $\{a \succ b, a \succsim b, a \prec b, a \precsim b, a \sim b, a \nsim b\}$. Furthermore, a proper subset $V \subseteq R$ of possible response patterns represents violations of a particular axiom, and the remaining set of response patterns $\sim V = R - V$ represent non-violations.

As a simple example of the concept of "possible response patterns," consider two gambles $a =$ (\$96,.95; \$0) and $b =$ (\$96,.95; \$24), where the form (\$$x$, p; \$$y$) refers to a binary a gamble where \$$x$ is won with probability p, and \$$y$ is won with probability $1-\text{p}$. Suppose that a given experiment generated N total responses, where for example, each of N subjects state a single preference between these gambles, or where a single subject stated a preference between these gambles over $N \geqslant 1$ repeated trials. The axiom of consequence monotonicity predicts that the strict preference relation $a \prec b$ must occur. From such an experiment, there may be $J=2$ possible response patterns over the two gambles a and b, characterized by $R_{ab}=(r_1 = a \succsim b,\ r_2 = a \prec b)$, where pattern $V_{ab}=r_1$ represents a violation of consequence monotonicity, and pattern $\sim V_{ab}=r_2$ represents a non-violation. In indifference responses are not allowed (forced-choice), then the possible response patterns are $R_{ab}=(r_1 = a \succ b, r_2 = a \prec b)$, where again, $V_{ab}=r_1$ and $\sim V_{ab}=r_2$. In a color-matching experiment, where $a \sim b$ signifies that light a is metametric to light b, the possible response patterns would be $R_{ab}=(r_1 = a \sim b,\ r_2 = a \nsim b)$.

It is possible to extend the concept of possible response patterns to handle more complex patterns of binary preferences. As an example, consider the transitivity axiom, which, for any set of three objects $\{a,b,c\}$, predicts that

$$a \succ b \quad \text{and} \quad b \succ c \quad \text{imply} \quad a \succ c \tag{1}$$

must hold. Again, suppose that an experiment generated N subject responses. For example, each of N subjects stated a strict preference between each and every pair of the three objects $\{a,b,c\}$, or where a single a subject stated a strict preference between each and every pair for N repeated trials. From such an experiment, there are $J=8$ possible response patterns, denoted by $R_{abc}=(r_1,r_2,r_3,r_4,r_5,r_6,r_7,r_8)$. Here, $r_1=\{a \succ b, b \succ c, a \succ c\}$, $r_2=\{a \succ b, b \prec c, a \succ c\}$, $r_3=\{a \prec b, b \succ c, a \succ c\}$, $r_4=\{a \succ b, b \prec c, a \prec c\}$, $r_5=\{a \prec b, b \succ c, a \prec c\}$, $r_6=\{a \succ b, b \succ c, a \prec c\}$, $r_7=\{a \prec b, b \prec c, a \succ c\}$, and $r_8=\{a \prec b, b \prec c, a \prec c\}$. Also, response patterns $V=\{r_6,r_7\}$ represent violations of transitivity (1), and the remaining patterns $\sim V=\{r_1,r_2,r_3,r_4,r_5,r_8\}$ represent non-violations.

Since each of the four types of axioms presented in Table 1 are characterized by a set of binary order relations, the current study focuses on the case where each r_j refers to a pattern of binary preferences. But indeed the same Bayesian framework proposed in this study can be easily be applied when each r_j characterizes a pattern of M-ary preferences, for $M \geqslant 2$ (e.g., Regenwetter, 1996).

2.2. Modeling a sequence of 0–1 random vectors

Now suppose that $\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N\}$ is a sequence of 0–1 random vectors that is observed from an experiment of size N, where any single observation $\mathbf{x}_i=(x_{i1},\ldots,x_{ij},\ldots,x_{iJ})$ is a 0–1 vector of length J, such that $x_{ij}=1$ if the ith response is response pattern r_j, and $x_{ij}=0$ otherwise, with $\sum_{j=1}^{J} x_{ij}=1$. Such a sequence may have been observed from N different subjects of a homogeneous group, or observed from the same subject over N repeated trials.

When modeling the joint probability $p(\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N)$ of the sequence, it can be reasonable to assume that the subscripts that identify the N individual observations $\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N\}$ are "uninformative," in the sense that the information that the $\mathbf{x}_i$'s provide are independent of the order they are collected. That is, the subscripts are *finitely exchangeable*, such that

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N)=p(\mathbf{x}_{\pi(1)},\ldots,\mathbf{x}_{\pi(i)},\ldots,\mathbf{x}_{\pi(N)}) \tag{2}$$

holds for any permutation π defined on the set $\{1,\ldots,i,\ldots,N\}$. An infinite sequence of random quantities $\{\mathbf{x}_1,\mathbf{x}_2,\ldots\}$ is *infinitely exchangeable* if every finite sequence of $\{\mathbf{x}_1,\mathbf{x}_2,\ldots\}$ is exchangeable.

The following representation theorem describes the form of the joint probability $p(\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N)$, when an observed sequence of 0–1 random vectors $\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N\}$ arises from an infinitely exchangeable sequence $\{\mathbf{x}_1,\mathbf{x}_2,\ldots\}$.

Representation Theorem for 0–1 Random Vectors (de Finetti, 1930, 1937/1964, Bernardo & Smith, 2002 Proposition 4.2, p. 176). *If* $\{\mathbf{x}_1,\mathbf{x}_2,\ldots\}$ *is an infinitely exchangeable sequence of* 0–1 *random vectors with probability measure P, then there exists a distribution function Q such that the joint mass function* $p(\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N)$ *for the finite sequence* $\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N\}$ *has the form*

$$p(\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N)=\int_{\Theta^*}\prod_{i=1}^{N}\prod_{j=1}^{J}\theta_j^{x_{ij}}\,dQ(\Theta), \tag{3}$$

where $\Theta^*=\{\Theta=(\theta_1,\ldots,\theta_j,\ldots,\theta_J);\ \theta_j\in[0,1],\ \sum_{j=1}^{J}\theta_j=1\}$, *and where*

$$Q(\Theta)=\lim_{N\to\infty}P[(\bar{x}_{1[N]}\leqslant\theta_1)\cup\cdots\cup(\bar{x}_{j[N]}\leqslant\theta_j)\cup\cdots\cup(\bar{x}_{J[N]}\leqslant\theta_J)]$$

for all $j=1,\ldots,J$, *with* $\bar{x}_{j[N]}=N^{-1}\sum_{i=1}^{N}x_{ij}$ *and* $\theta_j=\lim_{N\to\infty}\bar{x}_{j[N]}$.

For any exchangeable sequence of observations $\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N\}$, the resulting representation theorem has three implications. First, it is as if, conditional on Θ, the sequence $\{\mathbf{x}_1,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_N\}$ (observed from a homogeneous group of subjects or a single subject) is a random sample from a multinomial distribution with parameter Θ, according to a joint sampling distribution

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_N\mid\Theta)=\prod_{i=1}^{N}p(\mathbf{x}_i\mid\Theta)=\prod_{i=1}^{N}\prod_{j=1}^{J}\theta_j^{x_{ij}}. \tag{4}$$

Second, there *exists* a prior distribution Q which has to describe the initially available information about Θ

which labels the model. Hence, *a Bayesian approach is required* to model the joint probability $p(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N)$. And third, by the strong law of large numbers, $\theta_j = \lim_{N\to\infty} \bar{x}_{j[N]}$ for all $j = 1, \ldots, J$, so the distribution function Q may be interpreted as representing the prior beliefs about the limiting frequency of response pattern r_j.

Then from standard probability arguments involving Bayes' theorem, it follows that, after the outcome $\{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N\}$ of the experiment has occurred, the available information about Θ is described by its posterior density:

$$p(\Theta \mid \mathbf{x}_1, \ldots, \mathbf{x}_N) = \frac{\prod_{i=1}^{N} p(\mathbf{x}_i \mid \Theta) p(\Theta)}{\int \prod_{i=1}^{N} p(\mathbf{x}_i \mid \Theta) p(\Theta)\, d\Theta}, \tag{5}$$

where $p(\Theta)$ refers to the prior distribution of Θ in terms of a probability density function. However, since the representation theorem is an existence theorem, it does not specify the form of the prior distribution $p(\Theta)$ (or in more general terms, the form of the prior measure $Q(\Theta)$). Even so, it is possible to represent the prior density $p(\Theta)$ by the Dirichlet (Di) distribution:

$$\begin{aligned} p(\Theta) &= \left(\prod_{j=1}^{J} \Gamma(\tau_j)\right)^{-1} \Gamma\left(\sum_{j=1}^{J} \tau_j\right) \prod_{j=1}^{J} \theta_j^{\tau_j - 1} \\ &= \mathrm{Di}(\Theta \mid \tau_1, \ldots, \tau_j, \ldots, \tau_J), \end{aligned} \tag{6}$$

where $\Gamma(t>0) = \int_0^\infty q^{t-1} e^{-q}\, dq$ is the so-called Gamma function (Johnson & Kotz, 1970), and where $\tau_j > 0$ refers to the prior information about the number of observations of response pattern r_j. The Dirichlet prior distribution (6) is proper in the sense that it satisfies $\int p(\Theta)\, d\Theta \neq \infty$, and it is of the same family (conjugate to) the multinomial distribution, which means that the posterior distribution of Θ is available in closed form as a Dirichlet distribution:

$$\begin{aligned} p(\Theta \mid \mathbf{x}_1, \ldots, \mathbf{x}_N) &= \left(\prod_{j=1}^{J} \Gamma(n_j + \tau_j)\right)^{-1} \\ &\quad \times \Gamma\left(\sum_{j=1}^{J} (n_j + \tau_j)\right) \prod_{j=1}^{J} \theta_j^{n_j + \tau_j - 1} \\ &= \mathrm{Di}(\Theta \mid \tau_1 + n_1, \ldots, \tau_j + n_j, \ldots, \tau_J + n_J), \end{aligned} \tag{7}$$

where $n_j = \sum_{i=1}^{N} x_{ij}$ counts the number of responses equaling response pattern r_j from the N total observations. In fact, from a sequence of exchangeable observations $\{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N\}$, the vector $\mathbf{n} = (n_1, \ldots, n_J)$ (with $N = \sum_{j=1}^{J} n_j$) provides a sufficient statistic for the parameter Θ, thus $p(\Theta \mid \mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N) = p(\Theta \mid \mathbf{n})$. Furthermore, the posterior distribution (7) is proper because it satisfies $0 \leqslant \int p(\Theta \mid \mathbf{n})\, d\Theta \neq \infty$. This is important, because inferences based on an improper posterior distribution (where $\int p(\Theta \mid \mathbf{n})\, d\Theta = \infty$) are not guaranteed to be asymptotically consistent, in the sense that $|\hat{\Theta} - \Theta_0|$ does not necessarily converge to 0 as $N \to \infty$, where $\hat{\Theta}$ is the posterior mode and Θ_0 is the true population value of Θ (e.g., Gelman, Carlin, Stern, & Rubin, 2003, Section 4.3).

The posterior distribution (7) provides a basis for testing axioms on a single exchangeable sequence of data $\{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N\}$, while accounting for any prior information that may be available about the J possible response patterns $R = (r_1, \ldots, r_j, \ldots, r_J)$. However, when there is no prior information available, as is often the case in Bayesian analysis (e.g., Berger & Bernardo, 1992), it is reasonable to represent $p(\Theta)$ by a "non-subjective" prior that has little influence, resulting in a posterior distribution $p(\Theta \mid \mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N)$ that is mostly determined by the data.

One possible choice of a non-subjective prior distribution $p(\Theta)$ is the uniform distribution, specified by $\{\tau_1 = \tau_2 = \cdots = \tau_J = 1\}$ in (6) (and (7)). This prior distribution represents the belief that, before the sequence of data $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ is observed, each and every possible response pattern r_j is expected to occur with probability $1/J$. While the uniform prior distribution is proper (in the context of the Dirichlet prior distribution (6)), it is, unfortunately, not invariant over one-to-one transformations of Θ. It seems that such invariance is a minimum requirement for any method aimed at testing axioms that characterize theories of choice or measurement.

Alternatively, the reference prior distribution provides an appealing non-subjective choice for $p(\Theta)$, since it is invariant over the class of one-to-one transformations of Θ. This prior is also appealing because, with respect to all possible choices of proper prior distributions that could be defined on $p(\Theta)$, the reference prior is the choice of proper prior distribution that leads to a posterior distribution $p(\Theta \mid \mathbf{n}) \propto p(\mathbf{x}_1, \ldots, \mathbf{x}_N \mid \Theta) p(\Theta)$ that is *most* "data driven," i.e., that leads to a posterior $p(\Theta \mid \mathbf{n})$ that is mostly influenced by the sampling likelihood $\mathrm{p}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \mid \Theta)$, and least influenced by the prior $p(\Theta)$ (Bernardo, 1979). More precisely, the reference prior is the choice of prior distribution $p(\Theta)$ that maximizes the posterior information $\int p(\mathbf{n}) \int p(\Theta \mid \mathbf{n}) \log[p(\Theta \mid \mathbf{n})/p(\Theta)]\, d\mathbf{n}\, d\Theta$, where $p(\mathbf{n}) = \int p(\mathbf{n} \mid \Theta) p(\Theta)\, d\Theta$ is the marginal distribution, and $\int p(\Theta \mid \mathbf{n}) \log[p(\Theta \mid \mathbf{n})/p(\Theta)]\, d\Theta$ is the expected utility of the data $\mathbf{n}$ (Bernardo & Smith, 2002, Section 3.4.4).

The reference prior is specified according to how the analyst orders and groups the parameters $\theta_1, \ldots, \theta_J$ in terms of importance (Berger & Bernardo, 1992). For example, certain response patterns may be of special theoretical interest to the researcher, and thus s/he may treat a proper subset of the parameters $\theta_1, \ldots, \theta_J$ as the "parameters of interest," and treat the remaining parameters as "nuisance" parameters.[1] In this paper, for the task of axiom testing, all of the parameters will be treated as equally important, i.e., they all form a single group. Under this grouping, the reference prior is defined by the specification $\{\tau_1 = \tau_2 = \cdots = \tau_J = 1/2\}$. This prior is in fact equivalent to the Jeffrey's prior for the multinomial distribution (Jeffreys, 1961; Berger & Bernardo, 1992).

2.3. Axiom testing with the Bayes factor

In this study, an axiom is viewed as a hypothesis that implies a proper subset Ω of the total parameter space Θ^*, such that the possible values of Θ satisfy $\Theta \in \Omega \subseteq \Theta^*$. In particular, a deterministic axiom is viewed as "completely true" when $\underline{\theta} = \sum_{r_j \notin V} \theta_j = 1$, and so Ω is represented by this sum-constraint. The term $\underline{\theta}$ is the probability that a population of respondents yields observation level response patterns that are consistent with the theoretical-level order-relations that a specific axiom implies. In other words, the probability $\underline{\theta}$ indicates the degree to which an observed exchangeable sequence is consistent with the axiom, such that, in the limit as $\underline{\theta}$ approaches 1, it is increasingly certain that a deterministic axiom is true.

As is well-known (e.g., Bernardo & Smith, 2002, pp. 134–135), the posterior distribution of the sum $\underline{\theta} = \sum_{r_j \notin V} \theta_j$ can be written as

$$\begin{aligned} p(\underline{\theta} \mid s) &= \mathrm{Di}(\underline{\theta} \mid \tau_{\sim V} + s, \tau_V + N - s) \\ &= \mathrm{Be}(\underline{\theta} \mid \tau_{\sim V} + s, \tau_V + N - s), \\ &\propto \mathrm{Bin}(s \mid \underline{\theta})\, \mathrm{Be}(\underline{\theta} \mid \tau_{\sim V}, \tau_V), \end{aligned} \tag{8}$$

where Be refers to the Beta distribution, Bin denotes the binomial distribution, $s = \sum_{r_j \notin V} n_j$ is a sufficient statistic for $\underline{\theta}$ that counts the number of the N responses that do not "violate" a given axiom, $N - s$ is the number of responses that do "violate," and $\tau_{\sim V} = \sum_{r_j \notin V} \tau_j \geqslant 0$ and $\tau_V = \sum_{r_j \in V} \tau_j \geqslant 0$ represent the prior information that the axiom is non-violated and violated, respectively. Eq. (8) makes explicit that the Beta distribution is a special case of the Dirichlet distribution for $J = 2$. Furthermore, the posterior (8) is proper when and only when $\tau_{\sim V} + s > 0$ and $\tau_V + (N - s) > 0$.

Now suppose it was of interest to test the null hypothesis H_0 that a given axiom is satisfied with probability $\underline{\theta} \in [c_{\min}, c_{\max}] \subseteq [0, 1]$, against the general alternative hypothesis H_1: $\underline{\theta} \notin [c_{\min}, c_{\max}]$ that it is violated. Using conventional ideas of Bayesian inference (e.g., Carlin & Louis, 1996, pp. 47–54; Robert, 2001, p. 436; Bernardo & Smith, 2002, p. 436), the Bayes factor provides a device to compare the evidence in the data for H_0 against H_1. The Bayes factor is given by

$$\begin{aligned} &B(H_0\colon \underline{\theta} \in [c_{\min}, c_{\max}]) \\ &\quad = \frac{p(s \mid H_0)}{p(s \mid H_1)} = \frac{\int_{\underline{\theta} \in [c_{\min}, c_{\max}]} \mathrm{Bin}(s \mid \underline{\theta}) \mathrm{Be}(\underline{\theta})\, d\,\underline{\theta}}{\int_{\underline{\theta} \notin [c_{\min}, c_{\max}]} \mathrm{Bin}(s \mid \underline{\theta}) \mathrm{Be}(\underline{\theta})\, d\,\underline{\theta}} \\ &\quad = \frac{p(H_0 \mid s)/p(H_1 \mid s)}{p(H_0)/p(H_1)}. \end{aligned} \tag{9}$$

In the second term of the Bayes factor (9), $p(s \mid H_0)$ is the likelihood of the observed number of axiom non-violations s, given hypothesis H_0: $\underline{\theta} \in [c_{\min}, c_{\max}]$, and $p(s \mid H_1)$ is that likelihood given the alternative hypothesis H_1: $\underline{\theta} \notin [c_{\min}, c_{\max}]$. The third term of (9) shows that the Bayes factor is also the ratio of integrated likelihoods, where for example, $\mathrm{Bin}(s \mid \underline{\theta})$ denotes the binomial likelihood of the observed number of axiom non-violations s, integrated over the prior distribution $\mathrm{Be}(\underline{\theta})$ within the parameter space $\underline{\theta} \in [c_{\min}, c_{\max}]$ of the null hypothesis H_0.

Finally, in the last term of (9), the posterior odds ratio $p(H_0 \mid s)/p(H_1 \mid s)$ is the post-experimental evidence of H_0 after having observed the data s, and the prior odds-ratio $p(H_0)/p(H_1)$ is the pre-experimental evidence in favor of H_0. This last term in (9) shows that the Bayes factor $B(H_0 \in [c_{\min}, c_{\max}])$ is a ratio of these two odds, that measures how much the data observation s has increased ($B(H_0 \in [c_{\min}, c_{\max}]) > 1$) or decreased ($B(H_0 \in [c_{\min}, c_{\max}]) < 1$) the odds of H_0 relative to H_1, from prior (odds) to posterior (odds). The Bayes factor result of $B(H_0 \in [c_{\min}, c_{\max}]) > 1$ ($B(H_0 \in [c_{\min}, c_{\max}]) < 1$, respectively) means that there is more evidence in favor of H_0 versus H_1 (more evidence in favor of H_1 versus H_0, respectively). In fact, $\log(B(H_0 \in [c_{\min}, c_{\max}]))$ is the weight of evidence for H_0 (Good, 1950, 1985).

The Bayes factor (e.g., (9)) has properties that are necessary for coherent statistical tests. For example, it is invariant over all one-to-one transformations of the data (e.g., s) and parameters (e.g., $\underline{\theta}$). Also, it satisfies symmetry $B_{12} = 1/B_{21}$ and transitivity $B_{13} = B_{12}B_{23}$, where B_{12} denotes the Bayes factor of a hypothesis 1 over a hypothesis 2 (e.g., Robert, 2001, p. 351; Berger & Pericchi, 1996, p. 119). Furthermore, although (9) is formulated in terms of the posterior and prior probability of H_0, and of H_1, when using the Bayes factor, one does not have to strictly assume that either "model" H_0 or H_1 is true. This is because the Bayes factor is the integrated likelihood ratio weight of one hypothesis over another (second term of (9)), and therefore can be seen as measuring the relative success of H_0 versus H_1 at predicting the data (e.g., s) (Kass & Raftery, 1995; Berger & Pericchi, 1996). In fact (Dawid, 1992; Hartigan, 1992), the predictive ideas of the Bayes factor are related to so-called prequential analysis (Dawid, 1984) and stochastic complexity (Rissanen, 1987).

[1] Berger and Bernardo (1992) present a general algorithm for deriving the reference prior for any number of ordered groups of multinomial parameters.

It is straightforward to directly compute the Bayes factor (9), in terms of the posterior probability $p(H_0 \mid s)$ and the prior probability $p(H_0)$ (e.g., Carlin & Louis, 1996, p. 32). In particular

$$p(H_0 \mid s) = p(\underline{\theta} < c_{\max} \mid s) - p(\underline{\theta} \leqslant c_{\min} \mid s), \quad (10)$$

where $p(\underline{\theta} \leqslant c_{\max})$ (and $p(\underline{\theta} \leqslant c_{\min})$) is calculated through the cumulative distribution function of the beta posterior in (8):

$$p(\underline{\theta} \leqslant c \mid s) = \frac{\Gamma(N + \tau_{\sim V} + \tau_V)}{\Gamma(s + \tau_{\sim V})\Gamma(N - s + \tau_V)} \times \int_0^c z^{s+\tau_{\sim V}}(1-z)^{N-s+\tau_V}\, dz. \quad (11)$$

The prior probability of H_0, which is $p(H_0) = p(\underline{\theta} \leqslant c_{\max}) - p(\underline{\theta} < c_{\min})$, is calculated through (11) after specifying $s = N = 0$. Furthermore, $p(H_1 \mid s) = 1 - p(H_0 \mid s)$ and $p(H_1) = 1 - p(H_0)$.

In this study, when testing a deterministic axiom on a set of data with the Bayes factor $B(H_0 \in [c_{\min}, c_{\max}])$, five null hypotheses are considered separately, namely: $H_0{:}\,\underline{\theta} \geqslant .999$, $H_0{:}\,\underline{\theta} \geqslant .99$, $H_0{:}\,\underline{\theta} \geqslant .95$, $H_0{:}\,\underline{\theta} \geqslant .75$, and $H_0{:}\,\underline{\theta} \geqslant .50$. Consistent with Jeffreys' (1961) interpretation, the result $B(H_0 \in [c_{\min}, c_{\max}]) < 1/10$ is interpreted as indicating "decisive" evidence in the data against a null hypothesis H_0. Now, recall that in the limit as $\underline{\theta}$ approaches 1, it is increasingly certain that a deterministic axiom is true. According to this idea, then, if the Bayes factor does not indicate decisive evidence against $H_0{:}\,\underline{\theta} \geqslant .999$, it is concluded that the data do not violate a given deterministic axiom. If however the Bayes factor rejects all hypotheses except $H_0{:}\,\underline{\theta} \geqslant .50$ (that is, if each of the four other stricter hypotheses obtained aBayes factor less than 1/10), it is then concluded that the data "weakly" supports the deterministic axiom.

Occasionally, it is of interest to test an axiom under some *point-null* hypothesis θ_0. For example, in experiments conducted by Cho and colleagues (Cho, Luce, & von Winterfeldt, 1994; Cho & Luce, 1995), it was assumed that each subject had $J = 2$ possible preference responses for any paired presentation of gambles a and b, characterized by the possible response patterns ($r_1 = a \succ b, r_2 = a \prec b$). In testing axioms having the form $a \sim b$, these authors considered the null hypothesis $H_0{:}\,\underline{\theta}_0 = 1/2$ against the alternative $H_1{:}\,\underline{\theta}_0 \neq 1/2$. Accordingly, the Bayes factor $B(H_0 \in [c_{\min}, c_{\max}])$ can be calculated under the null hypothesis $H_0{:}\,|\underline{\theta} - \underline{\theta}_0| \leqslant \varepsilon$ against the general alternative $H_1{:}\,|\underline{\theta} - \underline{\theta}_0| > \varepsilon$, where $\varepsilon = c_{\max} - c_{\min}$ represents a "small" interval $[c_{\min}, c_{\max}] \subseteq [0,1]$ centered around $\underline{\theta}_0$(Berger & Delampady, 1987, p. 320). (But of course, in the posterior $p(\underline{\theta} \mid s)$ used to calculate this Bayes factor, s does not represent the observed number of axiom non-violations, as both response patterns $r_1 = a \succ b$ and $r_2 = a \prec b$ violate an axiom with form $a \sim b$. Here, s represents the number of observations of response pattern $r_1 = a \succ b$, and thus $N - s$ would then be the number of observations of response pattern $r_2 = a \prec b$.)

2.4. Testing axioms on partially exchangeable sequences

Thus far, we have only considered testing an axiom on a data set of size N arising from a *single* exchangeable observed sequence of 0–1 random vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N\}$, each observation labeled by a single index $i = 1, 2, \ldots$. In other words, a sequence of observations was assumed to arise from a homogeneous group of subjects, or the same subject over multiple trials. Clearly, in many applications of axiom testing, the situation is more complicated than this. As an example (not meant to describe all possible examples), a sequence observed from an experiment may be represented by six different subscripts, such as

$$\{\mathbf{x}_{1mc124}, \mathbf{x}_{2md137}, \mathbf{x}_{3fb125}, \mathbf{x}_{3fb225}, \mathbf{x}_{4mb134}, \ldots, \mathbf{x}_{imb234}, \ldots, \mathbf{x}_{Nfb448}\}, \quad (12)$$

where for instance $\mathbf{x}_{3fb125}$ denotes the response of subject $i = 3$, that subject being a female subject (f) with education level b (e.g., of five possible education levels a, b, c, d, e), at time point 1 (e.g., of four total time points) under experimental condition 2 (e.g., of four possible experimental conditions), with respect to axiom 5 (e.g., of the eight total axioms that were tested in the experiment). Certainly, it is not reasonable to model the sequence in (12) as exchangeable, because the rates of axiom violation may depend on subject, gender, education level, experimental condition, time point, and which axiom is being tested.

Given such a possibility of dependence, it is often more realistic to assume exchangeability of a sequence of observations $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}\}$ obtained from *within* a homogeneous unit k, for multiple homogeneous units $k = 1, \ldots, m$. For example, it may be reasonable to assume exchangeability of observations $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}\}$ obtained from the kth homogeneous group of subjects characterized by a particular gender, education level, time point, experimental condition, and axiom. As another example, it may be reasonable to assume exchangeability of the observations $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}\}$ obtained from the kth homogeneous group, characterized a particular subject, within a particular experimental condition, and with respect to a particular axiom.

In order to deal with such complicated situations, it is necessary to adapt the basic form of the Representation Theorem to accommodate *partial exchangeability*. This adaptation involves modeling a total sequence, such as (12), as m independent exchangeable sequences $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m\}$, where k indexes an exchangeable sequence of N_k observations $\{\mathbf{x}_{1_k}, \ldots,$

$\mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}\}$ obtained from a particular homogeneous unit.

When modeling the joint probability $p(\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m)$ of m independent exchangeable sequences, it is assumed that the sequences satisfy *unrestricted exchangeability*, in the sense that

$$p(\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m) = p(\mathbf{x}_{\pi_k(1_k)}, \ldots, \mathbf{x}_{\pi_k(i_k)}, \ldots, \mathbf{x}_{\pi_k(N_k)};\ k = 1, \ldots, m) \quad (13)$$

holds for any choice of permutation π_k on $\{1_k, \ldots, i_k, \ldots, N_k\}$, for $k = 1, \ldots, m$. Infinite sequences of random quantities $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, ;\ k = 1, \ldots, m\}$ are *unrestricted infinitely exchangeable* if every finite sequence of $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \}$ is exchangeable, for $k = 1, \ldots, m$. If m sequences of 0–1 random vectors $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}; k = 1, \ldots, m\}$ arises from m unrestricted infinitely exchangeable sequences, then the following representation theorem describes the form of the joint probability: $p(\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m)$.

Representation Theorem for *m* sequences of 0–1 random vectors (Bernardo & Smith, 2002, pp. 211–216). *If* $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots;\ k = 1, \ldots, m\}$ *are unrestricted infinitely exchangeable sequences of* 0–1 *random vectors with probability measure P, then there exists a distribution function Q such that the joint mass function* $p(\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m)$ *for the m finite sequences* $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m\}$ *has the form*

$$p(\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m) = \int_{\Theta^*} \prod_{k=1}^{m} \prod_{i_k=1}^{N_k} \prod_{j_k=1}^{J_k} \theta_{j_k}^{x_{i_k j_k}}\, dQ(\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m) \quad (14)$$

with $\Theta^* = \Theta_1^* \times \cdots \times \Theta_m^*$, $\Theta_k^* = \{\Theta_k = (\theta_{1_k}, \ldots, \theta_{j_k}, \ldots, \theta_{J_k}); \theta_{j_k} \in [0,1], \sum_{j_k=1}^{J_k} \theta_{j_k} = 1\}$, *and* $Q(\Theta_k) = \lim_{N_k \to \infty} P[(\bar{x}_{1_k[N_k]} \leqslant \theta_{1_k}) \cup \cdots \cup (\bar{x}_{j_k[N_k]} \leqslant \theta_{j_k}) \cup \cdots \cup (\bar{x}_{J_k[N_k]} \leqslant \theta_{J_k})]$ *for all* $j_k = 1, \ldots, J_k$ *and* $k = 1, \ldots, m$, *with* $\bar{x}_{j_k[N_k]} = N_k^{-1} \sum_{i_k=1}^{N_k} x_{i_k j_k}$, $\theta_{j_k} = \lim_{N_k \to \infty} \bar{x}_{j_k[N_k]}$.

Accordingly, with the prior density $p(\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m)$ assumed to be the product of m Dirichlet densities:

$$p(\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m) = \prod_{k=1}^{m} p(\Theta_k) = \prod_{k=1}^{m} \mathrm{Di}(\Theta_k \mid \tau_{1_k}, \ldots, \tau_{J_k}), \quad (15)$$

the joint posterior distribution of $\Theta_1, \ldots, \Theta_k$, conditional on m observed independent exchangeable sequences $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m\}$, is

$$p(\Theta_1, \ldots, \Theta_m \mid \mathbf{x}_{1_k}, \ldots, \mathbf{x}_{N_k}; k = 1, \ldots, m) = \prod_{k=1}^{m} \mathrm{Di}(\Theta_k \mid \tau_{1_k} + n_1, \ldots, \tau_{j_k} + n_{j_k}, \ldots, \tau_{J_k} + n_{J_k}), \quad (16)$$

and it follows that the joint posterior distribution $p(\underline{\theta}_1, \ldots, \underline{\theta}_k, \ldots, \underline{\theta}_m \mid s_1, \ldots, s_m)$ is

$$p(\underline{\theta}_1, \ldots, \underline{\theta}_m \mid s_1, \ldots, s_m) = \prod_{k=1}^{m} \mathrm{Be}(\underline{\theta}_k \mid \tau_{\sim V_k} + s_k, \tau_{V_k} + N_k - s_k), \quad (17)$$

where $\underline{\theta}_k$ denotes the posterior probability that an axiom is satisfied in the kth exchangeable sequence of observations $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}\}$, and s_k refers to the number of axiom non-violations observed in that sequence.

Of course, a set of m observed independent exchangeable sequences $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m\}$ can give rise to a set of m Bayes factors $\{B(H_{0k}\colon \underline{\theta}_k \in [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}]);\ k = 1, \ldots, m\}$, where $B(H_{0k}\colon \underline{\theta}_k \in [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}])$ refers to a test of the null hypothesis $H_{0k}\colon \underline{\theta}_k \in [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}] \subseteq [0,1]$ against the alternative hypothesis $H_{1k}\colon \underline{\theta}_k \notin [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}]$ on the kth sequence of data. The parameter space $[\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}]$ of a null hypothesis H_{0k} may be chosen to vary over the different sequences $k = 1, \ldots, m$.

Since the Bayes factor is multiplicative, as supported by its coherency properties mentioned in the previous subsection (symmetry, transitivity), it is possible to compute the Bayes factor for the evidence of the null hypothesis $H_0\colon \{\underline{\theta}_k \in [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}];\ k = 1, \ldots, m\}$ against the alternative hypothesis $H_1\colon \{\underline{\theta}_k \notin [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}];\ \text{any } k\}$, by the product:

$$B(H_0) = \prod_{k=1}^{m} B(H_{0k}\colon \underline{\theta}_k \in [\mathrm{c}_{\min,k}, \mathrm{c}_{\max,k}]). \quad (18)$$

Eq. (18) provides a general and flexible method for testing axioms on data that typically arise from experiments. For example, it may be of interest to test the null hypothesis H_0 that the data satisfy a single axiom, over m observed independent exchangeable sequences, where each sequence k is from a particular homogeneous group. Here, the kth sequence may characterize the responses from subjects of a particular gender, education level, experimental condition, and/or time point, or alternatively, the kth sequence may reflect multiple trials within a single subject. As a second example, given m independent observed exchangeable sequences, it may be of interest to test the null hypothesis H_0 satisfy m different axioms, where each sequence k corresponds to a test of a particular axiom. Of course, the Bayes factor (18) can be extended to test axioms in a situation reflecting some combination of the two preceding examples. In summary, the Bayes factor (18) can be aggregated over any m observed exchangeable sequences, provided that these sequences are independent.

3. Axiom tests

This section presents example applications of the Bayes framework described in Section 2, through tests of 16 different axioms on real data, axioms that are central to the theories of choice and measurement. Each of the 16 axioms belongs to one of the four types of axioms, as classified in Table 1. All these axioms were tested using a statistical program the author developed in the S-PLUS (1995) language, and is provided in the Appendix.

Nine of the axioms, relating to theories of choice behavior, will be tested, and they are detailed in Table 2. See Luce (2000) for a full theoretical background on these axioms. Table 2 contains notation in the form ($\$a$, p; $\$b$, q; $\$c$, r), which denotes a gamble where $\$a$ is won with probability p, $\$b$ is won with probability q, and $\$c$ is won with probability $r = 1 - \mathrm{p} - \mathrm{q}$. Also, Table 2 also refers to U(g), denoting the utility of a gamble g, and also refers to $g \oplus h$, denoting the joint receipt of two gambles g and h. As shown, the choice axioms refer specifically to gambles having money consequences. However, these axioms apply to choice behavior in general. For example, Steingrimmson (2002) consider choice axioms (with non-monetary consequences) in the psychophysics domain.

Table 2
Nine axioms for theories of choice

Axiom name	Axiom
Consequence monotonicity	$\$a \succsim \$b \Leftrightarrow (\$a, \mathrm{p};\ \$c) \succsim (\$b, \mathrm{p};\ \$c)$
Segregation	$(\$a \oplus \$b, \mathrm{p};\ \$b) \sim (\$a, \mathrm{p};\ 0) \oplus \$b$, for $a, b > 0$
Duplex decomposition	$(\$a, \mathrm{p}; -\$b) \sim (\$a, \mathrm{p};\ 0) \oplus (\$0, \mathrm{p}; -\$b)$, for $a, b > 0$
Additivity of joint receipt	$\mathrm{U}(g \oplus h) = \mathrm{U}(g) + \mathrm{U}(h)$
Stochastic dominance	$\mathrm{pr}(\$a > t \mid g) \geqslant \mathrm{pr}(\$a > t \mid h) \forall t \Rightarrow g \sim h$ or $g \succ h$
Coalescing	$(\$a, \mathrm{p};\ \$a, \mathrm{q};\ \$b, \mathrm{r}) \sim (\$a, p + \mathrm{q};\ \$b, \mathrm{r})$
Lower cumulative independence	$(\$a, \mathrm{p};\ \$b, \mathrm{q};\ \$c, \mathrm{r}) \succ (\$a, \mathrm{p};\ \$b', \mathrm{q};\ \$c', \mathrm{r}) \Rightarrow (\$b', \mathrm{p};\ \$c, \mathrm{q} + \mathrm{r}) \succ (\$b', \mathrm{p} + \mathrm{q}; \$c', \mathrm{r})$
Upper cumulative independence	$(\$a', \mathrm{p};\ \$b', \mathrm{q};\ \$c', \mathrm{r}) \succ (\$a, \mathrm{p};\ \$b, \mathrm{q};\ \$c', \mathrm{r}) \Rightarrow (\$a', \mathrm{p};\ \$b', \mathrm{q} + \mathrm{r}) \succ (\$a, \mathrm{p} + \mathrm{q};\ \$b', \mathrm{r})$
Restricted branch independence	$(\$a, \mathrm{p};\ \$b, \mathrm{q};\ \$c, \mathrm{r}) \succ (\$a', \mathrm{p};\ \$b', \mathrm{q};\ \$c, \mathrm{r}) \Rightarrow (\$a, \mathrm{p};\ \$b, \mathrm{q};\ \$c', \mathrm{r}) \succ (\$a', \mathrm{p};\ \$b', \mathrm{q};\ \$c', \mathrm{r})$

Note: Notation such as ($\$a$, p; $\$b$, q; $\$c$, r) refers to a gamble where $\$a$ is won with probability p, $\$b$ is won with probability q, and $\$c$ is won with probability $\mathrm{r} = 1 - \mathrm{p} - \mathrm{q}$. Also, g and h denote two different gambles, U(g) refers to the utility of a gamble g, and $g \oplus h$ refers to the joint receipt of gambles g and h, and pr($\cdot$) refers to the cumulative distribution function defined on dollar amount.

Table 3
Seven axioms of measurement theory

Axiom name	Axiom
Transitivity	$a \succsim b$ and $b \succsim c \Rightarrow a \succsim c$
Quadruple condition	$P(a \succ b) \geqslant P(c \succ d) \Leftrightarrow P(a \succ c) \geqslant P(b \succ d)$
Bi-cancellation	$P(d \succ e) \geqslant P(a \succ b)$ and $P(e \succ f) \geqslant P(b \succ c) \Rightarrow P(d \succ f) \geqslant P(a \succ c)$
Order independence	$aw \succsim bw \Leftrightarrow ax \succsim bx$, for $\forall x \in A_2$
Double cancellation	$ax \succsim bw$ and $by \succsim cx \Rightarrow ay \succsim cw$
Triple cancellation	$ax \succsim bw$ and $by \succsim cx$ and $cz \succsim dy \Rightarrow az \succsim dw$
Thomsen condition	$ay \sim cx$ and $by \sim cw \Rightarrow aw \sim bx$

Note: The notation $a, b, c, d, e, f, \ldots$ refer to objects (e.g., stimuli), while $P(a \succ b)$ refers to the probability that object a is preferred to object b. Also, $aw \in A_1 \times A_2$ refers to an object having attribute $a \in A_1 = \{a, b, c, d, \ldots\}$ and attribute $w \in A_2 = \{w, x, y, z, \ldots\}$, where $A_1 \cap A_2 = \emptyset$.

Also, seven of the axioms of measurement theory will be tested, and they are detailed in Table 3. Krantz, Luce, Suppes, & Tversky (1971, Chap. 6), Falmagne (1985), and Michell (1990) provide excellent backgrounds about these axioms. As shown in Table 3, the transitivity, quadruple condition, and bi-cancellation axioms pertain to objects (e.g., stimuli) from a single set $A = \{a, b, c, \ldots\}$. In contrast, the axioms of conjoint measurement theory, namely order independence, double cancellation, triple cancellation, and the Thomsen condition, refer to objects arising from the product set $A_1 \times A_2$, where (in general) $A_1 = \{a, b, c, \ldots\}$ and $A_2 = \{w, x, y, \ldots\}$ are two distinct sets, and $aw \in A_1 \times A_2$ denotes an object with attribute a and attribute w.

The following subsections describe the tests of the 16 axioms, using the Bayesian framework presented earlier.

3.1. Tests of consequence monotonicity

As a simple illustration, the Bayes factor is applied first to test the consequence monotonicity axiom on data. The data arise from an experiment conducted by von Winterfeldt, Chung, Luce, & Cho (1997, Table 2, p. 413), where 31 subjects stated a preference between two gambles, ($96, .95; $0) and ($96, .95; $24). Since in the second consequent of the two gambles, $24 > $0, then subjects should prefer the second gamble to the first, according to the axiom. The results of the experiment showed that $s = 19$ of the $N = 31$ subjects preferred the second gamble to the first, as predicted by consequence monotonicity. In the experiment, a subject was considered to prefer the second gamble to the first gamble, if and only if the subject judged the second gamble to be worth a higher sum of money than the first, i.e., had a higher "judged certainty equivalent."

Given observation $s = 19$ and $N = 31$, is there enough evidence in the data to support that the axiom is violated? To answer this question, it is assumed that the 31 observations in the sequence $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{31}\}$ are exchangeable, a random sample that depends on a multinomial distribution with parameter vector Θ. Also consider $J = 2$ possible response patterns, namely $r_1 = \{(\$96,.95;\ \$0) \succsim (\$96,.95;\ \$24)\}$ and $r_2 = \{(\$96,.95;\ \$0) \prec (\$96,.95;\ \$24)\}$, where only the second response pattern conforms to consequence monotonicity, thus $V = r_1$ and $\sim V = r_2$. In correspondence, the parameter vector is $\Theta = (\theta_1, \theta_2 = 1 - \theta_1)$, with $\theta_2 = \underline{\theta}$ and $\theta_1 = 1 - \underline{\theta}$, where $\underline{\theta}$ is the probability that consequence monotonicity is satisfied. Also, by specifying $\tau_1 = \tau_2 = 1/2$, the reference prior distribution is assumed over those possible response patterns. Simply, $\tau_1 = \tau_V = 1/2$ is the prior on the response pattern that represents a violation of the axiom, and $\tau_2 = \tau_{\sim V} = 1/2$ is the prior on the response pattern that represents a non-violation. It turns out that the data ($s = 19,\ N = 31$) "weakly" support consequence monotonicity, as the Bayes factor concludes rejection of $H_0{:}\,\underline{\theta} \geqslant .999$, $H_0{:}\,\underline{\theta} \geqslant .99$, $H_0{:}\,\underline{\theta} \geqslant .95$, and $H_0{:}\,\underline{\theta} \geqslant .75$, but does not reject the hypothesis $H_0{:}\,\underline{\theta} \geqslant .50$. In particular, $B[H_0{:}\,\underline{\theta} \geqslant .5] = 8.62$, with $p(H_0 \mid s) = .896$ and $p(H_0 \mid s) = .5$.

Table 4 shows the observed number s of the $N = 31$ responses not violating consequence monotonicity, for each of 15 different pairs of gambles $\{k = 1, \ldots, m = 15\}$ studied by von Winterfeldt et al. (1997), including the gamble pair just considered. It is assumed that for every gamble pair k, the 31 observations in the sequence $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{31_k}\}$ are exchangeable, a random sample from a multinomial distribution with parameter vector $\Theta_k = (\theta_{1_k}, \theta_{2_k} = (1 - \theta_{1_k}))$, with $\theta_{2_k} = \underline{\theta}_k$, and $\theta_{1_k} = (1 - \underline{\theta}_k)$. One column of Table 4 presents the results of the Bayes factor under the reference prior distribution $\{\tau_{1_k} = \tau_{V_k} = 1/2, \tau_{2_k} = \tau_{\sim V_k} = 1/2\}$ for all $k = 1, \ldots, 15$. As shown, one gamble pair does not reject $H_0{:}\,\underline{\theta}_k \geqslant .99$, two pairs do not reject $H_0{:}\,\underline{\theta}_k \geqslant .95$, two pairs do not reject $H_0{:}\,\underline{\theta}_k \geqslant .90$, four pairs fail to reject $H_0{:}\,\underline{\theta}_k \geqslant .75$, and the remaining six pairs fail to reject $H_0{:}\,\underline{\theta}_k \geqslant .50$. (Recall that when a null hypothesis, say $H_0{:}\,\underline{\theta}_k \geqslant .90$, is "not rejected", it is implied that the Bayes factor is at least 1/10 for all weaker hypotheses $\{H_0{:}\,\underline{\theta}_k \geqslant .75, H_0{:}\,\underline{\theta}_k \geqslant .50\}$, and less than 1/10 for all stricter hypotheses $\{H_0{:}\,\underline{\theta}_k \geqslant .95, H_0{:}\,\underline{\theta}_k \geqslant .99, H_0{:}\,\underline{\theta}_k \geqslant .999\}$).

Table 4 presents Bayes factors calculated under two other prior distributions that may be considered reasonable, as specified by $\{\tau_{1_k} = \tau_{V_k} = 1, \tau_{2_k} = \tau_{\sim V_k} = 4\}$ and

Table 4
Number of $N = 31$ subjects not violating consequence monotonicity (data from von Winterfeldt et al., 1997, Table 2), and the Bayes factors support of that axiom over different priors $\{\tau_1 = \tau_V, \tau_2 = \tau_{\sim V}\}$ (for clarity, subscript k is suppressed)

Gamble pair	Non-violations (s)	Null H_0: $\underline{\theta} \geqslant c_{min}$ that was not rejected (Bayes factor)		
		$\tau_V = 1/2$, $\tau_{\sim V} = 1/2$	$\tau_V = 1$, $\tau_{\sim V} = 4$	$\tau_V = 1$, $\tau_{\sim V} = 8$
1. ($96, .05; $0) vs. ($96, .05; $6)	23	⩾.75 (1.6)	⩾.75 (.51)	⩾.75 (.22)
2. ($96, .20; $0) vs. ($96, .20; $6)	15	⩾.50 (.75)	⩾.50 (.11)	None
3. ($96, .50; $0) vs. ($96, .50; $6)	17	⩾.50 (2.4)	⩾.50 (.36)	None
4. ($96, .80; $0) vs. ($96, .80; $6)	18	⩾.50 (4.4)	⩾.50 (.69)	⩾.50 (.14)
5. ($96, .95; $0) vs. ($96, .95; $6)	21	⩾.75 (.42)	⩾.75 (.15)	⩾.50 (2.3)
6. ($96, .05; $6) vs. ($96, .05; $24)	29	⩾.99 (.18)	⩾.99 (.13)	⩾.95 (.88)
7. ($96, .20; $6) vs. ($96, .20; $24)	27	⩾.95 (.22)	⩾.95 (.15)	⩾.90 (2.3)
8. ($96, .50; $6) vs. ($96, .50; $24)	21	⩾.75 (.42)	⩾.75 (.15)	⩾.50 (2.3)
9. ($96, .80; $6) vs. ($96, .80; $24)	18	⩾.50 (4.4)	⩾.50 (.69)	⩾.50 (.14)
10. ($96, .95; $6) vs. ($96, .95; $24)	19	⩾.50 (8.6)	⩾.50 (1.42)	⩾.50 (.327)
11. ($96, .05; $0) vs. ($96, .05; $24)	26	⩾.90 (.58)	⩾.90 (.29)	⩾.75 (.18)
12. ($96, .20; $0) vs. ($96, .20; $24)	24	⩾.75 (3.1)	⩾.75 (.97)	⩾.75 (.43)
13. ($96, .50; $0) vs. ($96, .50; $24)	25	⩾.90 (.21)	⩾.90 (.11)	⩾.75 (.89)
14. ($96, .80; $0) vs. ($96, .80; $24)	16	⩾.50 (1.3)	⩾.50 (.20)	None
15. ($96, .95; $0) vs. ($96, .95; $24)	20	⩾.95 (.20)	⩾.50 (3.19)	⩾.50 (.82)
Bayes factors calculated over multiple gamble pairs				
Bayes factor over gambles 1–5 (under H_0: $\underline{\theta} \geqslant .5$)		109181	16.20	.01
Bayes factor over gambles 5–10 (under H_0: $\underline{\theta} \geqslant .5$)		3.9×10^{15}	1.1×10^{12}	7.0×10^{9}
Bayes factor over gambles 11–15 (under H_0: $\underline{\theta} \geqslant .5$)		3.2×10^{12}	1.1×10^{9}	.01
Bayes factor for all 15 gambles (under H_0: $\underline{\theta} \geqslant .5$)		1.4×10^{33}	2.2×10^{22}	4.8×10^{14}

$\{\tau_{1_k} = \tau_{V_k} = 1, \tau_{2_k} = \tau_{\sim V_k} = 8\}$ for all $k = 1, \ldots, 15$. Each of these two priors represent the belief that, for any of the 15 gamble pairs, a subject is more inclined to prefer the second gamble over the first gamble. The results of Table 4 indicate that there is less evidence of consequence monotonicity, as there is more prior weight on the second gamble in a pair. Under prior $\{\tau_{1_k} = \tau_{V_k} = 1, \tau_{2_k} = \tau_{\sim V_k} = 8\}$, three gamble pairs violate the axiom.

It may be reasonable to assume independence between the five exchangeable sequences of 31 observations $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{31_k};\ k = 1, \ldots, 5\}$ arising from gamble pairs 1–5. Independence may also hold between the five exchangeable sequences of 31 observations $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{31_k};\ k = 6, \ldots, 10\}$ arising from each of the gamble pairs 6–10, and between the five exchangeable sequences $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{31_k};\ k = 11, \ldots, 15\}$ arising from each of the gamble pairs 11–15. However, it does not seem realistic to assume that all $m = 15$ exchangeable sequences of observations $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, ; k = 1, \ldots, m = 15\}$ are *independent*, because each of the 15 gamble pairs shares a gamble with some other gamble pair. According to the assumptions of independence that can be made, Table 4 presents the results of the Bayes factor aggregated over gamble pairs 1–5, to provide a test of the null hypothesis $H_0: \{\underline{\theta}_k \geqslant .5;\ k = 1, \ldots, 5\}$. Likewise, the Bayes factor was aggregated over gamble pairs 6–10 under the null hypothesis $H_0: \{\underline{\theta}_k \geqslant .5;\ k = 6, \ldots, 10\}$, and aggregated over gamble pairs 11–15 under the null hypothesis$H_0: \{\underline{\theta}_k \geqslant .5;\ k = 11, \ldots, 15\}$. With respect to priors $\{\tau_{1_k} = \tau_{V_k} = 1/2, \tau_{2_k} = \tau_{\sim V_k} = 1/2\}$ and $\{\tau_{1_k} = \tau_{V_k} = 1, \tau_{2_k} = \tau_{\sim V_k} = 4\}$, these aggregate Bayes factors indicate that there is very strong evidence that the data satisfy consequence monotonicity, within each of the three sets of gamble pairs 1–5, 6–10, and 11–15. But with respect to prior $\{\tau_{1_k} = \tau_{V_k} = 1, \tau_{2_k} = \tau_{\sim V_k} = 8\}$, the Bayes factor indicates that each of the set of gamble pairs 1–5 and 11–15 violate consequence monotonicity.

At the bottom of Table 4, it is seen that the Bayes factor, aggregated over all 15 pairs of gambles, indicates that the data, as a whole, satisfy consequence monotonicity, for any of the three choices of prior distribution. Though these aggregate Bayes factors should be interpreted with some caution. Since the 15 gamble pairs share gambles, there may be dependence in the choice responses among the subjects between those pairs, while the Bayes factor should only be aggregated over independent exchangeable sequences of data.

A reasonable way to address this response dependence is to increase the number of possible response patterns from $J = 2$ (for the moment, the subscript k is suppressed). For example, notice that gamble pairs 1, 6, and 11 share gambles (the same is true for gamble pairs 2,7,12, for pairs 3,8,13, or pairs 4,9,14, and for pairs 5,10,15). Instead of representing each of gamble pairs 1, 6, and 11 with $J = 2$ possible response patterns, these three pairs may be combined to form $J = 8$ possible response patterns, $R_{abc} = (r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8)$. Here, $r_1 = \{a \succsim b, b \succsim c, a \succsim c\}$, $r_2 = \{a \succsim b, b \succsim c, a \prec c\}$, $r_3 = \{a \prec b, b \succsim c, a \succsim c\}$, $r_4 = \{a \prec b, b \succsim c, a \prec c\}$, $r_5 = \{a \succsim b, b \prec c, a \succsim c\}$, $r_6 = \{a \succsim b, b \prec c, a \prec c\}$, $r_7 = \{a \prec b, b \prec c, a \succsim c\}$, and $r_8 = \{a \prec b, b \prec c, a \prec c\}$, where a, b denote the two gambles from pair 1, gambles b, c are from pair 6, and gambles a, c are from pair 11. From R_{abc} response pattern $\sim V = r_8$ represents a non-violation of consequence monotonicity, while each of the remaining patterns $V = \{r_1, r_2, r_3, r_4, r_5, r_6, r_7\}$ do represent violations. Now, with the Bayes factor, it is possible to test consequence monotonicity on the data frequencies $\mathbf{n} = (n_1, \ldots, n_8)$, and $s = n_8$, $(N - s) = n_1 + n_2 + n_3 + n_4 + n_5 + n_6 + n_7$. This approach to testing addresses any response dependency that may exist between gamble pairs 1, 6, and 11, because the possible response patterns in R_{abc} is constructed so as to combine the three pairs that share gambles. Unfortunately, this analysis could not be performed, since the paper by von Winterfeldt et al. (1997, Table 2, p. 413) presents the data as choice frequencies for each of the 15 gamble pairs separately, and not as choice frequencies in the form $\mathbf{n} = (n_1, \ldots, n_8)$, with respect to a vector of possible response patterns R_{abc}.

3.2. Tests of segregation, duplex decomposition, and additivity of joint receipt

Table 5 presents the choice frequencies of eighteen gamble pairs arising from subjects' judged certainty equivalents of each of the gambles, as reported in a study conducted by Cho et al. (1994, p. 938, Table 1). In Table 5, the frequencies are presented such that $A < B$ refer to an event where a subject judged the second gamble in a pair to have a higher certainty equivalent, $A > B$ refers to an event where the first gamble was judged higher. Finally, $A = B$ refers to an event where a subject judged both gambles in the pair to have the same certainty equivalent. In the following analysis, it is assumed that for each gamble pair $k = 1, \ldots, m = 18$, the 31 observations $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{31_k}\}$ are an exchangeable, random sample from a multinomial distribution with parameter Θ_k.

The data presented in Table 5 correspond to 6 tests of each of the axioms of segregation, duplex decomposition, and additivity of joint receipt. Recall that each of these three axioms is stated in the form $a \sim b$. So for each of the 18 gamble pairs $k = 1, \ldots, m = 18$, the vector of possible response patterns are considered as $R_{ab,k} = (r_{1_k} = a_k \succ b_k, r_{2_k} = a_k \prec b_k)$, where n_{1_k} and n_{2_k} are the choice frequencies of r_{1_k} and r_{2_k} respectively. Also, as done in Cho et al. (1994), the number of tied choices ($A = B$) is divided by two, and this number is added into each of the frequencies n_{1_k} and n_{2_k}.

Table 5
Number of subjects whose judged certainty equivalents (CEs) for one type of gamble were higher than, equal to, or lower than those for the other type of gamble, and Bayes factor values (data frequencies of the gamble pairs are from Table 1 of Cho et al., 1994)

Gamble pairs, segregation test	$A \geqslant B$	$A = B$	$A < B$	Bayes factor
(\$166, .2; \$70) vs. (\$96, .2; \$0)⊕ \$70	52	7	32	1.31
(\$166, .5; \$70) vs. (\$96, .5; \$0)⊕ \$70	53	8	30	.64
(\$166, .9; \$70) vs. (\$96, .9; \$0)⊕ \$70	37	29	34	12.01
(−\$166, .2; \$70) vs. (−\$96, .2; \$0)⊕ −\$70	11	9	36	.029
(−\$166, .2; \$70) vs. (−\$96, .5; \$0)⊕ −\$70	21	11	24	8.69
(−\$166, .2; \$70) vs. (−\$96, .9; \$0)⊕ −\$70	21	4	31	3.84
				Overall Bayes factor for segregation = 9.67
Gamble pairs, duplex decomposition test				
(\$96, .2; −\$40) vs. (96, .2; \$0)⊕ (0, .2; −\$40)	33	13	45	5.42
(\$96, .5; −\$40) vs. (96, .5; \$0)⊕(0, .5; −\$40)	36	19	36	11.99
(\$96, .9; −\$40) vs. (96, .9; \$0)⊕ (0, .9; −\$40)	44	10	37	9.16
(\$96, .2; −\$160) vs. (96, .2; \$0)⊕ (0, .2; −\$160)	31	20	40	7.68
(\$96, .5; −\$160) vs. (96, .5; \$0)⊕ (0, .5; −\$160)	35	20	36	11.92
(\$96, .9; −\$160) vs. (96, .9; \$0)⊕ (0, .9; −\$160)	34	19	38	10.98
				Overall Bayes factor for duplex decomposition = 598215.8
Gamble pairs, additivity of joint receipt test				
(\$96, .2,\$0)⊕ (\$70) vs. (\$96, .2,\$0) + (\$70)	26	17	48	.82
(\$96, .5,\$0)⊕ (\$70) vs. (\$96, .5,\$0) + (\$70)	25	17	49	.49
(\$96, .9,\$0)⊕ (\$70) vs. (\$96, .9,\$0) + (\$70)	15	17	59	.0002
(−\$96, .2,\$0)⊕ (−\$70) vs. (−\$96, .2,\$0)+ (−\$70)	27	5	24	8.69
(−\$96, .5,\$0)⊕ (−\$70) vs. (−\$96, .5,\$0)+ (−\$70)	29	8	19	3.84
(−\$96, .9,\$0)⊕ (−\$70) vs. (−\$96, .9,\$0)+ (−\$70)	34	12	10	.05
				Overall Bayes factor for additivity of joint receipt = .0001

Overall Bayes factor for segregation, duplex decomposition, and additivity of joint receipt = 650.13

Note: Each result of the Bayes factor refers to the test of H_0: $|\underline{\theta}-\underline{\theta}_0|\leqslant\varepsilon$ against the alternative H_1: $|\underline{\theta}-\underline{\theta}_0|>\varepsilon$, where $\underline{\theta}_0 = .5$ and $\varepsilon = 2\times10^{-6}$ (k subscript suppressed).

According to the possible response patterns specified, the parameter vector for each gamble pair k has the form $\Theta_k = (\theta_{1_k}, \theta_{2_k} = (1-\theta_{1_k}))$, with $\theta_{1_k} = \underline{\theta}_k$ treated as the probability of choice r_{1_k}, and $\theta_{2_k} = (1-\underline{\theta}_k)$ the probability of choice r_{2_k}. Also, the reference prior is specified with $\tau_{1_k} = \tau_{2_k} = 1/2$. Since each of the axioms is of the form $a \sim b$, the Bayes factor, for each pair k, is computed under the null hypothesis H_{0k}: $|\underline{\theta}_k - .5| \leqslant \varepsilon$, against the general alternative H_1: $|\underline{\theta}_k - .5| > \varepsilon$, where ε represents a "small" interval centered at the point null-hypothesis $\underline{\theta}_{0k} = .5$. This interval is $\varepsilon = 2\times10^{-6}$, but can be chosen to be arbitrarily small.

Table 5 shows the results of the Bayes factor, which state that the data violate segregation in gamble pair $k = 4$, none of six gamble pairs violate duplex decomposition, and gamble pairs $k = 15$ and $k = 18$ violate additivity of joint receipt. Cho et al. (1994, Table 1) calculated the χ^2 sign test under the point-null hypothesis H_0: $\theta_{0k} = .5$, and also reported that gamble pairs $k = 4, 15, 18$ violated their respective axioms (i.e., rejected the point-null hypothesis). In fact, among the 18 pairs of gambles, these three gambles had the highest χ^2 values under the null hypothesis, while here, these three gambles had the lowest Bayes factors. However, these authors reported that gamble pairs $k = 1, 2, 4, 13, 14, 18$ also violated their respective axioms. These latter χ^2 results may reflect the well-known fact that the p-value, arising from the χ^2 test, tends to overstate the evidence against the null hypothesis, especially for large sample sizes (Berger & Delampady, 1987, Section 4.6; Delampady & Berger, 1990).

It may be realistic to assume independence between the 18 exchangeable sequences of observations $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k}; k = 1, \ldots, m = 18\}$ arising from all gamble pairs 1–18, because none of the gamble pairs share any gambles between them. This means that it is possible to aggregate the Bayes factor over all the 18 exchangeable sequences, in any manner.

Table 5, shows that the Bayes factor was aggregated over gamble pairs 1–6, to test the null hypothesis H_0: $\{|\underline{\theta}_k - .5| \leqslant \varepsilon;\ k = 1, \ldots, 6\}$, and it is concluded that all these pairs, together, satisfy segregation. Also, under the null hypothesis $H_0\colon \{|\underline{\theta}_k - .5| \leqslant \varepsilon;\ k = 7, \ldots, 12\}$, the Bayes factor indicates that there is enormous evidence in the data that the set of gamble pairs 7–12 satisfy duplex decomposition. However, under the null hypothesis $H_0\colon \{|\underline{\theta}_k - .5| \leqslant \varepsilon; k = 13, \ldots, 18\}$, the Bayes factor indicates that the set of gamble pairs 13–18 violate additivity of joint receipt. Of course, it is possible to test whether there is evidence that the data satisfy all three axioms, by calculating the Bayes factor under the null hypothesis $H_0\colon \{|\underline{\theta}_k - .5| \leqslant \varepsilon; k = 1, \ldots, 18\}$. As the bottom of Table 5 shows, all eighteen pairs of gambles satisfy segregation, duplex decomposition, and additivity of joint receipt, together. Though this global result should not overshadow the fact that gamble pairs 13–18 violate additivity of joint receipt.

3.3. *Tests of many other axioms of theories of choice and measurement*

We now proceed to present several other data sets, in Tables 6–10, assumed to contain eleven exchangeable sequences $k = 1, \ldots, m = 11$. These sequences provide tests of the 12 remaining axioms, outlined in Tables 2 and 3. Each of these data sets (sequences) is presented as frequencies in a mutidimensional contingency table, where an un-bolded frequency represents the number of observations of a response pattern r_{j_k} that characterizes a violation of a specific axiom (that is, $r_{j_k} \in V_k$). A bold frequency represents the number of observations of a particular response pattern r_{j_k} that represents a non-violation of a specific axiom (that is, $r_{j_k} \in\sim V_k$).

Table 6 presents data from Birnbaum's (1999) experiment, as four separate 2×2 contingency tables, each indicating the observed frequencies of preference-responses made over two pairs of gambles (each single gamble is described in that table). The frequency data were generated from a lab-based experiment where 124 subjects stated preferences between pairs of gambles, over two repeated trials. Thus there were 248 responses for each gamble pair (aside from missing data arising from occasional subject non-response). The objective of Birnbaum's experiment was to test the several axioms of choice behavior, including stochastic dominance, coalescing, lower cumulative independence, upper cumulative independence, and restricted branch independence.

Tables 7 and 8 refer to a data set of 14 cancer patients who indicated, on each pair of a set of symptoms $A = \{a, b, c, d, e, f\}$ (described in the tables), the symptom they prefer to learn more about. This data set is available in http://umanitoba.ca/centres/mchp/concept/

Table 6
Data sets for tests of stochastic dominance, coalescing, cumulative independence, and branch independence, on gamble preference data (Birnbaum, 1999, Tables 4–7). Frequencies in bold represent instances where the axiom is not violated (for stochastic dominance and coalescing: bold and underline)

Gambles of two paired comparisons — Choice frequencies

(a) *Data for tests of stochastic dominance and coalescing*

$a = (\$12,\ .05;\ \$14,\ .05;\ \$96,\ .90)$,
$a' = (\$12,\ .10; \$90,\ .05;\ \$96,\ .85)$,
$b = (\$12,\ .05;\ \$14,\ .05;\ \$96,\ .05;\ \$96,\ .85)$,
$b' = (\$12,\ .05;\ \$12,\ .05;\ \$90,\ .05;\ \$96,\ .85)$.

	$a \succ a'$	$a \prec a'$
$b \succ b'$	**<u>54</u>**	155
$b \prec b'$	12	**24**

(b) *Data for the test of lower cumulative independence*

$c = (\$2,\ .80;\ \$40,\ .10;\ \$44,\ .10)$,
$d = (\$2,\ .80;\ \$10,\ .10;\ \$98,\ .10)$,
$c'' = (\$10,\ .80;\ \$44,\ .20)$,
$d'' = (\$10,\ .90;\ \$98,\ .10)$.

	$c \succ d$	$c \prec d$
$c'' \succ d''$	**46**	**29**
$c'' \prec d''$	57	**114**

(c) *Data for the test of upper cumulative independence*

$c' = (\$40,\ .10;\ ,\$44 .10,;\ \$110 .80)$,
$d' = (\$10, .10\ ;\ \$98,\ .10;\ \$110,\ .80)$,
$c''' = (\$40,\ .20;\ \$98,\ .80)$,
$d''' = (\$10,\ .10;\ \$98,\ .90)$.

	$c' \succ d'$	$c' \prec d'$
$c''' \succ d'''$	**57**	106
$c''' \prec d'''$	**9**	**73**

(d) *Data for the test of restricted branch independence*

$c = (\$2,\ .80;\ \$40,\ .10;\ \$44,\ .10)$
$d = (\$2,\ .80;\ \$10,\ .10;\ \$98,\ .10)$.
$c' = (\$40,\ .10;\ \$44,\ .10;\ \$110,\ .80)$
$d' = (\$10,\ .10;\ \$98,\ .10;\ \$110,\ .80)$.

	$c \succ d$	$c \prec d$
$c' \succ d'$	**46**	19
$c' \prec d'$	55	**124**

Table 7
Data for tests of transitivity and the quadruple condition on the symptoms data set

Items of paired comparisons | Choice frequencies

(a) *Data for a test of transitivity*
a = Insomnia
b = Bowel
d = Fatigue

		$b \succ d$	$b \prec d$
$a \succ d$	$a \succ b$	**1**	**2**
	$a \prec b$	**2**	2
$a \prec d$	$a \succ b$	0	**2**
	$a \prec b$	**5**	**0**

(b) *Data for a test of the quadruple condition*
a = Insomnia
b = Bowel
e = Breathing
f = Pain

			$e \succ f$	$e \prec f$
$b \succ f$	$a \succ e$	$a \succ b$	**2**	**2**
		$a \prec b$	**2**	**4**
	$a \prec e$	$a \succ b$	0	0
		$a \prec b$	**1**	2
$b \prec f$	$a \succ e$	$a \succ b$	**0**	**1**
		$a \prec b$	0	**0**
	$a \prec e$	$a \succ b$	**0**	**0**
		$a \prec b$	**0**	**0**

The data frequencies in bold indicate instances where the corresponding axiom is not violated.

Table 8
Data for the test of bi-cancellation on the symptoms data set

Choice frequencies

					$d \succ e$	$d \prec e$
$d \succ f$	$a \succ c$	$e \succ f$	$b \succ c$	$a \succ b$	**1**	**0**
				$a \prec b$	**1**	**0**
			$b \prec c$	$a \succ b$	**1**	**0**
				$a \prec b$	**1**	**0**
		$e \prec f$	$b \succ c$	$a \succ b$	**0**	**0**
				$a \prec b$	**2**	**0**
			$b \prec c$	$a \succ b$	**2**	**0**
				$a \prec b$	**0**	**0**
	$a \prec c$	$e \succ f$	$b \succ c$	$a \succ b$	**0**	**0**
				$a \prec b$	**1**	**0**
			$b \prec c$	$a \succ b$	**0**	**0**
				$a \prec b$	**0**	**0**
		$e \prec f$	$b \succ c$	$a \succ b$	**0**	**0**
				$a \prec b$	**1**	**0**
			$b \prec c$	$a \succ b$	**0**	**0**
				$a \prec b$	**1**	**0**
$d \prec f$	$a \succ c$	$e \succ f$	$b \succ c$	$a \succ b$	0	**0**
				$a \prec b$	0	0
			$b \prec c$	$a \succ b$	0	**0**
				$a \prec b$	0	0
		$e \prec f$	$b \succ c$	$a \succ b$	**0**	**0**
				$a \prec b$	**1**	**1**
			$b \prec c$	$a \succ b$	0	**0**
				$a \prec b$	0	0
	$a \prec c$	$e \succ f$	$b \succ c$	$a \succ b$	**0**	**0**
				$a \prec b$	**0**	**0**
			$b \prec c$	$a \succ b$	**0**	**0**
				$a \prec b$	**0**	**0**
		$e \prec f$	$b \succ c$	$a \succ b$	**0**	**0**
				$a \prec b$	**0**	**0**
			$b \prec c$	$a \succ b$	**0**	**0**
				$a \prec b$	**0**	**0**

The data frequencies in bold indicate instances where the axiom is not violated.
Note: The items for paired comparisons are: a = insomnia, b = bowel, c = fatigue, d = appetite, e = breathing, f = pain.

thurstone/programming.html. The data presented in Table 7a provide a test of transitivity arising from three paired comparisons of symptoms a, b, and d, and the data in Table 7b provide a test of the quadruple condition using four paired comparisons of symptoms a, b, e, and f. (Recall that the quadruple condition is defined in Table 3 in terms of choice probabilities.) Also, the data in Table 8 provide a test of the bi-cancellation condition arising from six paired comparisons of all six symptoms.

Tables 9 and 10 refer to data arising from a memory experiment (obtained from W. H. Batchelder and Jarad Smith, University of California at Irvine). The data in these two tables provide tests of axioms of conjoint measurement. In this experiment, each of 109 subjects were instructed to study a set of words for a short time period, and after the study period, and then to recall the words over four separate repeated trials.[2] The tables refer to the set of trials as $A_1 = \{a = \text{Trial 1}, b = \text{Trial 2}, c = \text{Trial 3}, d = \text{Trial 4}\}$, and four words, defined by the set $A_2 = \{w = \text{Word 1}, x = \text{Word 2}, y = \text{Word 3}, z = \text{Word 4}\}$. For any member of the product set $aw \in A_1 \times A_2$, in terms of the data, a single observation is coded $aw = 1$ when a subject successfully recalls word a on trial w, and $aw = 0$ otherwise. Therefore, for any pair $aw, bx \in A_1 \times A_2$, a relation $aw \succsim bx$ in Tables 9 and 10a refers to a data observation where the response on aw is greater than or equal to the response on bx, and $aw \prec bx$ when response bx is greater than response aw. Similarly, a relation $aw \sim bx$ in Table 10b refers to a data observation where the response on aw is equal to the response on bx, and "not $aw \sim bx$" is an observation where they are not equal.

The sixth column of Table 11 (labeled "1-GROUP") presents the results of the Bayes factor, for tests of each of 12 axioms of decision and measurement, on the data sets described in Tables 6–10. In this column, it is assumed that for each test of the 12 axioms, the sequence of observations $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k}\}$ arising from a group of N_k subjects, are exchangeable, a random sample from a multinomial distribution that depends on a parameter $\Theta_k = (\theta_{1_k}, \theta_{2_k}, \ldots, \theta_{j_k} \ldots, \theta_{J_k})$, with $\underline{\theta}_k = \sum_{r_{j_k} \notin V_k} \theta_{j_k}$ and $1 - \underline{\theta}_k = \sum_{r_{j_k} \in V_k} \theta_{j_k}$. Moreover, the Bayes factor is considered under five different null hypotheses, namely: $H_0{:}\,\underline{\theta}_k \geqslant .999$, $H_0{:}\,\underline{\theta}_k \geqslant .99$, $H_0{:}\,\underline{\theta}_k \geqslant .95$, $H_0{:}\,\underline{\theta}_k \geqslant .75$, and $H_0{:}\,\underline{\theta}_k \geqslant .50$. Also, recall that each axiom is assigned a prior specification $\{\tau_{\sim V_k}, \tau_{V_k}\}$

[2] Of course, in this data example; the solvability axiom of conjoint measurement is violated unless the four levels of the "Trials" factor are equally spaced (see Krantz et al., 1971, Chapters. 1 and 6).

Table 9
Data for tests of independence and double cancellation on the (simulated) memory experiment data

Pattern frequencies

(a) *Data for a test of ordinal independence*

			$bx \succsim ax$	$bx \prec ax$
$bz \succsim az$	$by \succsim ay$	$bw \succsim aw$	**74**	4
		$bw \prec aw$	9	0
	$by \prec ay$	$bw \succsim aw$	8	1
		$bw \prec aw$	1	0
$bz \prec az$	$by \succsim ay$	$bw \succsim aw$	9	1
		$bw \prec aw$	1	0
	$by \prec ay$	$bw \succsim aw$	0	1
		$bw \prec aw$	0	**0**

(b) *Data for a test of double cancellation*

			$by \succsim cx$	$by \prec cx$
text	$ay \succsim cw$	$ax \succsim bw$	**17**	**13**
		$ax \prec bw$	**11**	5
	$ay \prec cw$	$ax \succsim bw$	16	**16**
		$ax \prec bw$	**18**	**13**

The data frequencies in bold indicate instances where the corresponding axiom is not violated.
Note: The objects of conjoint measurement are trials: $A_1 = \{a = \text{Trial 1}, b = \text{Trial 2}, c = \text{Trial 3}, d = \text{Trial 4}\}$, and words: $A_2 = \{a = \text{Word 1}, b = \text{Word 2}, c = \text{Word 3}, d = \text{Word 4}\}$.

Table 10
Data for tests of triple cancellation and the Thomsen condition on the (simulated) memory-experiment data

Pattern frequencies

(a) *Data for a test of triple cancellation*

			$by \succsim cx$	$by \prec cx$
$az \succsim dw$	$cz \succsim dy$	$ax \succsim bw$	**13**	**8**
		$ax \prec bw$	**10**	**2**
	$cz \prec dy$	$ax \succsim bw$	**4**	**2**
		$ax \prec bw$	**2**	3
$az \prec dw$	$cz \succsim dy$	$ax \succsim bw$	12	**18**
		$ax \prec bw$	**13**	**11**
	$cz \prec dy$	$ax \succsim bw$	**4**	**1**
		$ax \prec bw$	**4**	**2**

(b) *Data for a test of the Thomsen condition*

		$by \sim cw$	not $by \sim cw$
$aw \sim bx$	$ay \sim cx$	**17**	**13**
	not $ay \sim cx$	**11**	5
not $aw \sim bx$	$ay \sim cx$	16	**16**
	not $ay \sim cx$	**18**	**13**

The data frequencies in bold indicate instances where the corresponding axiom is not violated.
Note: The objects of conjoint measurement are trials: $A_1 = \{a = \text{Trial 1}, b = \text{Trial 2}, c = \text{Trial 3}, d = \text{Trial 4}\}$, and words: $A_2 = \{a = \text{Word 1}, b = \text{Word 2}, c = \text{Word 3}, d = \text{Word 4}\}$.

Table 11
Tests of 12 axioms with the Bayes factor (for clarity, k subscript is suppressed)

Axiom	Prior $\tau_{\sim V}$	Prior τ_V	s	N	Null H_0: $\underline{\theta} \geqslant c_{\min}$ that was not rejected (Bayes factor)	
					1-Group	Individual
Stochastic dominance and coalescing	.5	1.5	54	245	Axiom violated	Axiom violated
Coalescing only	1	1	78	245	Axiom violated	Axiom violated
Lower cumulative ind.	1.5	.5	189	246	$\geqslant .75$ (2.0)	$\geqslant .75$ (2.7×10^6)
Upper cumulative ind.	1.5	.5	139	245	$\geqslant .50$ (14.6)	Axiom violated
Restricted branch independence	1	1	170	244	$\geqslant .50$ (3.9×10^9)	$\geqslant .75$ (6.8×10^{10})
Transitivity	3	1	12	14	$\geqslant .95$ (.32)	$\geqslant .95$ (.31)
Quadruple condition	6	2	12	14	$\geqslant .95$ (.41)	$\geqslant .95$ (.39)
Bi-cancellation	27.5	4.5	13	13	$\geqslant .999$ (5.19)	$\geqslant .999$ (7.10)
Order independence	1	7	74	109	$\geqslant .75$ (78.7)	$\geqslant .95$ (7.2×10^{19})
Double cancellation	3	1	88	109	$\geqslant .75$ (9.0)	$\geqslant .75$ (1.2×10^5)
Triple cancellation	7	1	94	109	$\geqslant .90$ (.12)	$\geqslant .75$ (4.5×10^4)
Thomsen condition	3	1	88	109	$\geqslant .75$ (9.0)	$\geqslant .75$ (1.2×10^5)

Note: "Prior $\tau_{\sim V}$" refers to the prior $\tau_{\sim V} = \sum_{r_j \notin V} \tau_j$ that an axiom is satisfied, and "Prior τ_V" refers to the prior $\tau_V = \sum_{r_j \in V} \tau_j$ that the axiom is violated (assuming the reference prior over all J possible response patterns).

according to the number of response patterns the axiom contains in $\sim V_k$, and the number of response blocks it contains in V_k, with respect to the reference prior specified over all relation blocks, $\{\tau_{1_k} = \cdots = \tau_{J_k} = 1/2\}$. So for example, Table 9b shows that 2 of the eight response patterns violate double cancellation, and the 6 remaining response patterns do not violate. Given the reference prior $\{\tau_{1_k} = \cdots = \tau_{J_k} = 1/2\}$, the prior specification $\{\tau_{\sim V_k}, \tau_{V_k}\}$ for double cancellation is derived as $\tau_{\sim V_k} = \sum_{r_{j_k} \notin V_k} \tau_{j_k} = 3$ and $\tau_{V_k} = \sum_{r_{j_k} \in V_k} \tau_{j_k} = 1$.

The sixth column of Table 11 (labeled "1-GROUP") shows that the data violate both stochastic dominance and coalescing, and coalescing alone; in each case, $H_0{:}\,\underline{\theta} \geqslant .50$ was rejected, with the Bayes factor being near zero. But other data sets appear to give "strong" support for several axioms, such as transitivity ($H_0{:}\,\underline{\theta}_k \geqslant .95$ is not rejected by the Bayes factor), the quadruple condition ($H_0{:}\,\underline{\theta}_k \geqslant .95$ not rejected), bi-cancellation ($H_0{:}\,\underline{\theta}_k \geqslant .999$ not rejected), and triple cancellation ($H_0{:}\,\underline{\theta}_k \geqslant .90$ is not rejected). Also, the data give "moderate" support for lower cumulative independence, order independence, double cancellation ($H_0{:}\,\underline{\theta}_k \geqslant .75$ is not rejected), and "weak" support for upper cumulative independence and restricted branch independence ($H_0{:}\,\underline{\theta}_k \geqslant .50$ is not rejected).

The seventh column of Table 11 (labeled "INDIVIDUAL") shows the results of the Bayes factor, calculated under a different assumption of exchangeability that accounts for individual (subject) differences in response behavior. In particular, exchangeability is assumed within each subject (for the five axioms considered from Birnbaum's (1999) experiment, exchangeability is assumed within each trial as well). More precisely, within each of the 12 axioms, and within each subject *now* indexed by k, the single response of the kth subject $\{\mathbf{x}_{1_k}\}$ is viewed as a random sample that depends on a multinomial parameter vector Θ_k, assuming $\{\mathbf{x}_{1_k}\}$ is from an exchangeable sequence $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, ...\}$. (We could imagine unobserved multiple trials within that subject.) One advantage to this assumption of exchangeability is that, since responses are nested within each subject, the assumption of independence between choice pairs is avoided.

Then within each axiom, over the parameter vectors $\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m$, one vector per subject k, the Bayes factor in support of the null hypothesis H_0 that the axiom is satisfied is simply calculated through Eq. (18) by $B(H_0) = \prod_{k=1}^{m} B(H_{0k} : \underline{\theta}_k \in [c_{\min,k}, c_{\max,k}])$. Compared with the Bayes factors in the sixth column of Table 11 (labeled "1-Group"), the Bayes factors of the seventh column (labeled "Individual") that account for individual (subject) differences, provide four different conclusions. In particular, order independence and restricted branch independence are more strongly satisfied, triple cancellation is more weakly satisfied, and upper cumulative independence is violated.

4. Conclusions, connections, and generalizations

The Bayesian framework proposed in this paper provides a practical basis testing data fit to all types of deterministic axioms choice and measurement, for a wide range of experimental setups. The focus of the current study was the exchangeable multinomial model:

$$p(\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k}; k = 1, \ldots, m) = \int_{\Theta \in \Omega \subset \Theta^*} \prod_{k=1}^{m} \prod_{i_k=1}^{N_k} \prod_{j_k=1}^{J_k} \theta_{j_k}^{x_{i_k j_k}} \, dQ (\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m), \tag{19}$$

where $\Theta^* = \Theta_1^* \times \cdots \times \Theta_m^*$, $\Theta_k^* = \{\Theta_k = (\theta_{1_k}, \ldots, \theta_{j_k}, \ldots, \theta_{J_k}); \theta_{j_k} \in [0,1], \sum_{j_k=1}^{J_k} \theta_{j_k} = 1\}$, and Ω is a subset of Θ^* implied by a specific deterministic axiom. The model in (19) provides a Bayesian framework for the so-called "Random Relations" approach (Regenwetter, 1996) to representing judgements and preferences. This approach entails assigning a probability measure on possible response patterns of M-ary preference relations (it is straightforward to extend (19) such that each possible response pattern consists of a set of M-ary relations, for $M \geqslant 2$). Furthermore, there are two other general ways to represent judgements and preferences. They are the random utility approach (which assumes that a real-valued utility random variable is associated with each choice alternative), and the random relations approach (which assumes that the set of choice alternatives are represented as a utility function, being a random sample from an urn of such utility functions). It is proven that, under reasonable conditions, that if any one of the three representations hold, then so do each of the others (Regenwetter & Marley, 2001).

Model (19), in terms of J_k, generalizes two other statistical models that were proposed for axiom testing. First, the binomial model, a special-case of (19) assuming $J_k = 2$, has been extensively investigated by Karabatsos (2001), Karabatsos & Ullrich (2003) and Karabatsos & Sheu (2004) in Bayesian tests of conjoint measurement axioms on binomial data. These axioms included independence, double cancellation, joint independence, distributive cancellation, and dual-distributive cancellation (see Karabatsos, 2005, for a review). In this context, the subset Ω represented order-restrictions on the binomial parameters, such as $\{\theta_{1_k} \leqslant \cdots \leqslant \theta_{j_k} \leqslant \cdots \leqslant \theta_{J_k}\}$. Second, Iverson & Falmagne's (1985) and Iverson (1991), in a classical statistical (non-Bayesian) framework, also proposed a binomial model with $J_k = 2$, where the probability of m exchangeable sequences is conditioned on the maximum-likelihood estimate of the vector $\Theta = (\Theta_1, \ldots, \Theta_m)$:

$$p(\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k}; k = 1, \ldots, m) = \sup_{\Theta \in \Omega \subset \Theta^*} \left(\prod_{k=1}^{m} \prod_{i_k=1}^{N_k} \prod_{j_k=1}^{J_k=2} \theta_{j_k}^{x_{i_k j_k}} \right). \tag{20}$$

Here, $\Theta = (\Theta_1, \ldots, \Theta_m)$ is constrained to lie within a subset Ω, representing the parametric order-constraints implied by any *probabilistic* axiom, such as weak stochastic transitivity:

$$P(a \succ b) \geqslant 1/2 \quad \text{and } P(b \succ c) \geqslant 1/2 \text{ imply}$$
$$P(a \succ c) \geqslant 1/2, \text{ for all } a, b, c \in A, \tag{21}$$

or the quadruple condition (see Table 3).

For four reasons, model (19) that handles the case of $J_k \geqslant 2$ provides a significant improvement over the two models that strictly assume $J_k = 2$. First, in the task of axiom testing, model (19) does not require a deterministic axiom to be reformulated into a probabilistic axiom (e.g., does not require deterministic transitivity to be reformulated as weak stochastic transitivity). Second, the two models that assume $J_k = 2$ tend to demand computationally-intensive methods for axiom testing. In particular, they either require Markov chain Monte Carlo simulation of the posterior distribution of the order-constrained parameters $p(\Theta \mid \mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m)$ (Gelfand, Smith, & Lee, 1992; Chib & Greenberg, 1995), or require the derivation of highly complex optimization methods to find the maximum likelihood estimate of Θ under the restriction $\Theta \in \Omega \subseteq \Theta^*$ (Iverson & Falmagne, 1985). In contrast, as demonstrated in Section 3, model (19) can be implemented to test any of the four types of axioms on data, over many different types of experiments, using a Bayes factor that can be directly calculated. Third, the non-Bayesian model (20) does not account for uncertainty in the parameter vector $\Theta = (\Theta_1, \ldots, \Theta_m)$, since it predicts that the probability of the m exchangeable sequences $\{\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, ;\ k = 1, \ldots, m\}$ is conditioned on the maximum likelihood estimate of Θ subject to the constraint $\Theta \in \Omega \subset \Theta^*$. In contrast, the Bayesian model (19) accounts for the uncertainty in model parameters, by predicting that the m exchangeable sequences is based on the integration of the multinomial likelihood over the prior distribution $Q(\Theta_1, \ldots, \Theta_m)$.

Fourth, model (19) easily accommodates any dependence that exists between paired comparisons that share objects. The two other models that imply $J_k = 2$ assume that the vector of possible response patterns is $R_k = (r_{1_k} = a \succ b,\ r_{2_k} = a \prec b)$, for the kth object pair $a, b \in A$, and for all distinct pairs of objects $k = 1, \ldots, m$. Obviously, this specification of possible response patterns leads to the parameterization $\Theta_k = (\theta_{1_k} = P(a \succ b),\ \theta_{2_k} = (1 - \theta_{1_k}) = P(a \prec b))$ for each exchangeable sequence $k = 1, \ldots, m$. But this assumption of $J_k = 2$ may be unrealistic for all experimental setups, because this leads to the assumption that any subject's choices between two objects are independent over the m paired comparisons (see (19)). On other words, it leads to the assumption that, for any two object pairs $k, k' \in \{1, \ldots, m\}$, a subject's preference between a and b within a pair k is *independent* from his/her preference between another b and c in a different pair k', even though the same object b appears in both paired comparisons. Of course, one can design an experiment with conditions where subjects are induced to make independent responses between all m unique pairs of objects $a, b \in A$, and/or by presenting subjects m pairs such that any two pairs k, $k' \in \{1, \ldots, m\}$, one pair does not consist of the same objects as another pair k'. But even so, for the analysis of choice data, it seems more satisfying to consider a model that accommodates any dependence that exists between paired comparisons that share objects. As mentioned, (19) does provide such a model, and it seems to provide a method for testing axioms on data arising from a larger class of experiments, not just data arising from experiments where subjects provided m independent paired-comparison judgements. Model (19) is able to accommodate $J_k \geqslant 2$ possible response patterns of binary preference relations $\{r_{1_k}, r_{2_k}, \ldots, r_{J_k}\}$, so that within a response pattern, the binary preference relations are over paired comparisons that share objects. As a simple example: $R_{abc,k} = (r_{1_k} = \{a \succ b, b \succ c\},\ r_{2_k} = \{a \prec b, b \succ c\},\ r_{3_k} = \{a \succ b, b \prec c\},\ r_{4_k} = \{a \prec b, b \prec c\})$, where object b is shared. Recall that Section 3 provided many other examples of this approach, in tests of restricted branch independence, lower and upper cumulative independence, transitivity, double cancellation, and so forth.

To conclude, (19) can be generalized to a hierarchical model, by not only by treating each of the m sequences $\{\mathbf{x}_{1_k}, \ldots, \mathbf{x}_{i_k}, \ldots, \mathbf{x}_{N_k}; k = 1, \ldots, m\}$ as exchangeable, but also by treating the parameters $\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m$ as an exchangeable, random sample from a distribution that depends on a hyper- parameter ϕ with prior distribution $\Pi(\phi)$. This hierarchical model is given by

$$p(\mathbf{x}_{1_k}, \mathbf{x}_{2_k}, \ldots, \mathbf{x}_{N_k};\ k = 1, \ldots, m) = \int_{\Theta \in \Omega \subset \Theta^*} \prod_{k=1}^{m} \prod_{i_k=1}^{N_k} \prod_{j_k=1}^{J_k} \theta_{j_k}^{x_{i_k j_k}}\, dQ$$
$$(\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m \mid \phi), Q(\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m \mid \phi) = \prod_{k=1}^{m} \phi(\Theta_k)$$
$$\Pi(\phi). \tag{22}$$

Model (22) offers a flexible framework that allows for any dependence that may exist between the parameters $\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m$, whereas model (19) strictly assume that these parameters are independent. Independence may not be realistically assumed when the parameters $\Theta_1, \ldots, \Theta_k, \ldots, \Theta_m$ can be judged as being "similar." Such judgements of similarity may be reasonable when the same subject is observed under m different conditions, or when m different subjects are observed under a particular experimental condition (e.g., time point). Judgements of similarity may also be reasonable when m different experiments are involved in the same study, or when m different paired comparisons of an experiment share choice objects between them. Hierarchical models of the form (22) deserve study in future research on methods of axiom testing. In fact, if $\Pi(\cdot)$ is specified by a Dirichlet process prior (Ferguson, 1973; Lo, 1984) or by a Póeya Tree prior (Mauldin, Sudderth, &

Williams, 1992; Lavine, 1992, 1994), then it is possible with (22) to generalize axiom testing to be based entirely on a non-parametric framework of statistical inference.

Acknowledgments

This research was supported by National Science Foundation Grant SES-0242030, and by Spencer Foundation Grant SG2001000020, the author as Principal Investigator on both grants. Thanks to John Miyamoto, the Action Editor, two anonymous referees, and Duncan Luce, for many detailed comments on previous versions of this paper. I also thank (in alphabetical order) William Batchelder, Michael Birnbaum, and Michel Regenwetter, for several insightful conversations.

Appendix

An S-Plus program for computing the Bayes factor, for testing a deterministic axiom under an interval-null hypothesis.

```
# Computes the Bayes Factor for an interval null hypothesis H0
# In the Box below, enter the number of axiom non-violations,
# the total number of trials, priors Tau(~V) and Tau(V),
# the lower and upper bounds of the interval null hypothesis,
# and then type F10 and view the results.
########################################################################
s <<- c( 0, 1, 2, 3, 4, 5, 6, 7, 8, 9,10)  # s, number of axiom non-    #
                                           #                violations #
N <<- c(10,10,10,10,10,10,10,10,10,10,10)  # N, total number of trials  #
a <<- rep(.5,count.rows(s))                # Beta prior tau ~V         #
b <<- rep(.5,count.rows(s))                # Beta prior tau V          #
H0.LB  <<- .50       # Lower bound of null hypothesis H0                #
H0.UB  <<- 1         # Upper bound of null hypothesis H0                #
########################################################################
BF1           <<-    function(a,b,s,N,H0.LB,H0.UB){
Prior.H0      <-     pbeta(H0.UB,a,b) - pbeta(H0.LB,a,b)
Post.H0       <-     pbeta(H0.UB,a+s,b+N-s) - pbeta(H0.LB,a+s,b+N-s)
Bayes.Factor  <-     (Post.H0/(1-Post.H0)) / (Prior.H0/(1-Prior.H0))
report        <-     cbind(a,b,deparse.level=0)
report        <-     cbind(report,s,deparse.level=0)
report        <-     cbind(report,N,deparse.level=0)
report        <-     cbind(report,round(Prior.H0,5))
report        <-     cbind(report,round(Post.H0,5))
report        <-     cbind(report,round(Bayes.Factor,5),deparse.level=0)
report        <-     rbind(matrix(c("Pri a","Pri
   b","s","N","Pri.H0","Post.H0","BF H0"),1,7),report)
Bayes.Factor.Global <- prod(Bayes.Factor)
return(report,Bayes.Factor.Global)}
BF1(a,b,s,N,H0.LB,H0.UB)

#### EXAMPLE RESULTS OF THE DATA ANALYSIS ####
#### USING THE SPECIFICATIONS OF s, N, a, b, H0.LB, AND H0.UB, DESCRIBED
#### IN THE PROGRAM ABOVE

 [1,] Pri a  Pri b  s  N  Pri.H0  Post.H0  BF H0
 [2,]  0.5    0.5   0  10  0.5    0.00016  0.00016
 [3,]  0.5    0.5   1  10  0.5    0.00369  0.0037
 [4,]  0.5    0.5   2  10  0.5    0.02604  0.02673
 [5,]  0.5    0.5   3  10  0.5    0.10202  0.1136
 [6,]  0.5    0.5   4  10  0.5    0.26483  0.36022
 [7,]  0.5    0.5   5  10  0.5    0.5      1
 [8,]  0.5    0.5   6  10  0.5    0.73517  2.77605
 [9,]  0.5    0.5   7  10  0.5    0.89798  8.80243
[10,]  0.5    0.5   8  10  0.5    0.97396  37.40745
[11,]  0.5    0.5   9  10  0.5    0.99631  270.01077
[12,]  0.5    0.5   10 10  0.5    0.99984  6192.47951

$Bayes.Factor.Global: 1
```

```
#### OTHER EXAMPLE DATA SETS ####
#
# Data from Von Winterfeldt, Chung, Luce & Cho (1997, J Exp Psych:LMC,
# Table 2)
# for tests of consequence monotonicity
# s <<- c(23,15,17,18,21,29,27,21,18,19,26,24,25,16,20)
# N <<- 31
# a <<- rep(.5,count.rows(s))
# b <<- rep(.5,count.rows(s))
#
#
# Data from Cho,Luce,& von Winterfeldt (1994, J Exp Psych:HPP,Table 1)
# for tests of
# Segregation, Duplex Decomposition, and Additivity of Joint Receipt.
# s <<- c(55.5,57,51.5,15.5,26.5,23,39.5,45.5,49,41,45,43.5,34.5,33.5,
#                                                         23.5,29.5,33,40)
# N <<- c(  91,91, 100,  56,  56,56,  91,  91,91,91,91,  91,  91,  91,
#                                                           91,  56,56,56)
# a <<- rep(.5,count.rows(s))
# b <<- rep(.5,count.rows(s))
# H0.LB <<- .499999
# H0.UB <<- .500001
#
#
# Data for (1-GROUP) tests of
# Stochastic Dominance, Coalescing, Lower Cumulative Independence,
# Upper cumulative Independence, Restricted Branch Independence,
# Transitivity, Quadruple Condition, Bi-cancellation, Order
# Independence, Double Cancellation, Triple Cancellation, and the
# Thomsen Condition.
# s <<- c( 54,  78, 189, 139, 170, 12, 12,   13,  74,  88,  94,  88)
# N <<- c(245, 245, 246, 245, 244, 14, 14,   13, 109, 109, 109, 109)
# a <<- c( .5,   1, 1.5, 1.5,   1,  3,  6, 27.5,   1,   3,   7,   3)
# b <<- c(1.5,   1,  .5,  .5,   1,  1,  2,  4.5,   7,   1,   1,   1)
#
#
# Data for a test of Triple Cancellation, taking into account individual
# differences.
# s    <<- c(rep(1,94),rep(0,109-94))
# N    <<- c(rep(1,109))
# a    <<- rep(7,109)
# b    <<- rep(1,109)
```

References

Berger, J. O., & Bernardo, J. M. (1992). Ordered group reference priors with an application to the multinomial problem. *Biometrika*, *79*, 25–37.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*, 317–352.

Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109–122.

Bernardo, J. M. (1979). Reference posterior distributions for Bayes inference (with discussion). *Journal of the Royal Statistical Society, B*, *41*, 113–147.

Bernardo, J. M. (1996). The concept of Exchangeability and its applications. *Far East Journal of Mathematical Sciences*, *4*, 111–121.

Bernardo, J. M., & Smith, A. F. M. (2002). *Bayesian theory* (second reprint). New York: Wiley.

Birnbaum, M. (1999). Testing critical properties of decision making on the internet. *Psychological Science*, *10*, 399–407.

Carlin, B. P., & Louis, T. A. (1996). *Bayes and empirical bayes methods for data analysis* (first reprint). Boca Raton, FL: Chapman & Hall/CRC.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis–Hastings algorithm. *American Statistician*, *49*, 327–335.

Cho, Y.-H., & Luce, R. D. (1995). Tests of hypotheses about certainty equivalents and joint receipt of gambles. *Organizational Behavior and Human Decision Processes*, *64*, 229–248.

Cho, Y.-H., Luce, R. D., & von Winterfeldt, D. (1994). Tests of assumptions about the joint receipt of gambles in rank- and sign-dependent utility theory. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 931–943.

Dawid, A. P. (1984). Statistical theory, the prequential approach. *Journal of the Royal Statistical Society, Series A*, *147*, 178–292.

Dawid, A. P. (1992). Prequential analysis, stochastic complexity and Bayesian Inference (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics*, Vol. 4 (pp. 95–110). Oxford: Oxford University Press.

de Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. *Academia Nazionale dei Lincei*, *4*, 86–133.

de Finetti, B. (1937/1964). La prévision: ses lois logiques, ses sources subjectives. *Ann Inst. H. Poincaré*, 7, 1–68 (Reprinted in 1980 as 'Foresight; its logical laws, its subjective sources' In H. E. Kyburg, & H. E. Smokler (Eds.), *Studies in subjective probability* (pp. 93–158) New York: Dover.)

de Finetti, B. (1970/1974). *Theory of probability* (Vol. 1) (English translation by A. Machí and A. F. M. Smith as *Teoria delle Probabilitá* 1 in 1974). London: Wiley.

de Finetti, B. (1970/1975). *Theory of probability* (Vol. 2) (English translation by A. Machí and A. F. M. Smith as *Teoria delle Probabilitá* 2 in 1975). London: Wiley.

Delampady, M., & Berger, J. O. (1990). Lower bounds on Bayes factors for multinomial distributions with applications to χ^2 tests of fit. *Annals of Statistics*, *18*, 1295–1316.

Falmagne, J.-C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, *1*, 209–230.

Gelfand, A. E., Smith, A. F. M., & Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, *87*, 523–532.

Gelman, A., Carlin, J., Stern, H., & Rubin, D.B. (2003). *Bayesian data analysis* (second edition). Boca Raton, FL, Chapman & Hall/CRC.

Good, I. J. (1950). *Probability and the weighing of evidence*. London: Griffin.

Good, I. J. (1985). Weight of evidence: A brief survey (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics*, Vol. 2 (pp. 249–270). North-Holland: Amsterdam.

Hartigan, J. A. (1992). Locally uniform prior distributions. *Annals of Statistics*, *24*, 160–173.

Iverson, G. J. (1991). Probabilistic measurement theory. In J.-P. Doignon, & J.-C. Falmagne (Eds.), *Mathematical psychology: current developments* (pp. 134–135). New York: Springer.

Iverson, G. J., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences*, *10*, 131–153.

Jeffreys, H. (1961). *Theory of probability*. London: Oxford University Press.

Johnson, N. L., Kotz, S. (1970). *Continuous univariate distributions* (Vol. 2). Boston: Houghton-Mifflin.

Karabatsos, G. (2001). The Rasch model additive conjoint measurement and new models of probabilistic measurement theory. *Journal of Applied Measurement*, *2*, 389–423.

Karabatsos, G. (2005). Additivity tests. In B. Everitt, & Howell (Eds.), *Encyclopedia of behavioral statistics*. New York: Wiley.

Karabatsos, G., & Sheu, C.-F. (2004). Order-constrained Bayes inference for dichotomous non-parametric item-response theory. *Applied Psychological Measurement*, *28*, 110–125.

Karabatsos, G., & Ullrich, J. R. (2003). Enumerating and testing conjoint measurement models. *Mathematical Social Sciences*, *43*, 487–505 (Special issue on *Random utility theory and probabilistic measurement theory*, A. A. J. Marley, guest editor).

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement* (Vol. I). New York: Academic Press.

Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modeling. *Annals of Statistics*, *20*, 1222–1235.

Lavine, M. (1994). More aspects of Pólya tree distributions for statistical modeling. *Annals of Statistics*, *22*, 1161–1176.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates. *Annals of Statistics*, *12*, 351–357.

Luce, R. D. (2000). *Utility of gains and losses: measurement theoretical and experimental approaches*. Mahwah, NJ: Lawrence Erlbaum.

Mauldin, R. D., Sudderth, W. D., & Williams, S. C. (1992). Pólya trees and random distributions. *Annals of Statistics*, *20*, 1203–1221.

Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Regenwetter, M. (1996). Random utility representations of finite *m*-ary relations. *Journal of Mathematical Psychology*, *40*, 219–234.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology*, *45*, 864–912.

Rissanen, J. (1987). Stochastic complexity. *Journal of the Royal Statistical Society, Series B*, *49*, 223–239 and 252–265 (with discussion).

Robert, C. P. (2001). *The Bayesian choice*. New York: Springer.

S-PLUS (1995). S-*plus documentation*. Seattle: Statistical Sciences, Inc.

Steingrimmson, R. (2002). *Contributions to measuring three psychophysical attributes*: *Testing behavioral axioms for loudness*, *response time as an independent variable*, *and attentional intensity*. Ph.D. thesis, University of California-Irvine.

von Winterfeldt, D., Chung, N.-K., Luce, R. D., & Cho, Y. (1997). Tests of consequence monotonicity in decision making under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 406–426.