University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 381

Updating Stochastic Choice

Carlos Alós-Ferrer and Maximilian Mihm

March 2021

Updating Stochastic Choice

Carlos Alós-Ferrer
Zurich Center for Neuroeconomics (ZNE), Department of Economics, University of Zurich

Maximilian Mihm
Division of Social Science, New York University Abu Dhabi

This version: March 2021

When an economic agent makes a choice, stochastic models predicting those choices can be updated. The structural assumptions embedded in the prior model condition the updated one, to the extent that the same evidence produces different predictions even when previous ones were identical. We provide a general framework for models of stochastic choice allowing for arbitrary forms of (structural) updating and show that different models can be sharply separated by their structural properties, leading to axiomatic characterizations. Our framework encompasses Bayesian updating given beliefs over deterministic preferences (as implied by popular random utility models) and standard neuroeconomic models of choice, which update decision values in the brain through reinforcement learning.

KEYWORDS: Stochastic preferences, Bayesian learning, Logit choice, Reinforcement, Neuroeconomic theory.

1. Introduction

Consider the problem of developing a probabilistic model predicting the future choices of an agent given past, observable behavior. With new choice observations, any prior model of the agent's choices can be refined and updated. However, updating might take different forms depending on the structural assumptions of the initial model. For example, Bayesian updating may be natural in some cases, but not if the initial model is not formulated in terms of prior beliefs. What does updating mean in such a general context, and how can different updating rules be compared and characterized?

As a first example, consider an external observer who has a belief on the deterministic preference of an economic agent; that is, the observer's model is a probability distribution on the set of (strict) preferences, corresponding to a classical *random preference model* as in McFadden and Richter (1990), which is a cornerstone of many stochastic choice theories. When the agent makes a new choice, the observer can apply Bayes' rule to obtain a posterior distribution which is, in itself, a new model of stochastic choice within the same class. Equivalently, suppose that the observer's model describes a *population*, that is, the distribution describes the frequencies of different types, each characterized by a deterministic preference. With individual-level data, the observer applies Bayes' rule to *classify* agents, e.g. in terms of the most likely type in the corresponding posterior, or the average value of an individual-level preference parameter. Such classification problems underlie finite-mixture models (e.g., El-Gamal and Grether, 1995,

Carlos Alós-Ferrer: carlos.alos-ferrer@econ.uzh.ch
Maximilian Mihm: max.mihm@nyu.edu

We thank José Apesteguía, Ernst Fehr, Fabio Maccheroni, Jawwad Noor, and seminar participants at the University of Zurich for helpful comments. C.A.-F. gratefully acknowledges financial support from the Swiss National Science Foundation (SNF) under project 100014–179009.

Costa-Gomes et al., 2001, Bruhin et al., 2010, 2018). They also reflect current approaches to preference estimation, where individual heterogeneity is modeled through an estimated population distribution of preference parameters (the initial model), and individual-level parameters are updated from the individual choices (e.g., see Bellemare et al., 2008, Conte et al., 2011, Moffatt, 2015).

As a second example, consider a probabilistic choice model where a choice is made with a probability proportional to a transformation of an underlying value or utility, as in the classical model of Luce (1959), which is equivalent to logit choice, or as in more general models following Debreu (1958). The parameter of the model is not a prior distribution over types, but rather a collection of utilities, and "updating" should be formulated in terms of the latter. This example is particularly important in view of developments in neuroeconomics and decision neuroscience. Extensive evidence from these fields indicates that economic choices follow certain decision values computed in the brain (Schultz, 1998, Padoa-Schioppa and Assad, 2006, Ballesta et al., 2020), but those values are built on the basis of reinforcement learning (e.g., Holroyd and Coles, 2002, Daw and Tobler, 2014). Further, the mapping from decision values to choices is intrinsically stochastic (e.g. Shadlen and Kiani, 2013) to the extent that standard models in the literature always rely on logit or similar choice functions to derive choice frequencies from decision values.[1] Under this approach, given an actual, observable choice, the human brain will update the decision value through a reinforcement process, resulting in an updated internal model generating stochastic choices. The "observer" and the "agent" become proxies for different brain functions concerned with updating decision values and implementing choices, respectively.

As a third example, suppose that preferences are not stable over time, but are rather subject to change on the basis of past choices or anchors (Brehm, 1956, Ariely and Norton, 2008, Ariely, Loewenstein, and Prelec, 2003). This is a view frequently expressed in psychology, in particular in the field of cognitive dissonance (Festinger, 1957, Joule, 1986). Although seldom expressed formally, "choice-induced preference change" results in a formal object similar to the first two examples, for the observation of a choice leads to an update of the original model which does not necessarily respect Bayesian principles (as the "true" model changes after the new choice is made).[2]

In this paper, we develop a formal binary-choice framework capturing all these (and other) situations. A decision maker is described according to a *stochastic choice model* which predicts *a priori* stochastic choices that are updated as choices are observed *without violating the underlying structural assumptions of the model.* Thus, we consider datasets that include not only stochastic choices but rather history-dependent stochastic choices, where *histories* consist of chains of successive previous choice observations. A model rationalizing such datasets must explain not only the probabilities of choices, but also the structural properties by which those are updated. We show that the fundamental cases described above are particular (but important) examples of our framework, and provide full characterizations of the datasets (including histories of choices) that can be rationalized by a given stochastic choice model, which includes

[1] The Drift-Diffusion Model of Ratcliff (1978), widely used in cognitive science and neuroscience, predicts logit probabilities in binary choice. It can be seen as capturing neural processes which implement choices given decision values (Shadlen and Shohamy, 2016), or as an evidence accumulation process aiming to uncover underlying utilities (Fudenberg, Strack, and Strzalecki, 2018). See also Brocas and Carrillo (2012) for a related model.

[2] Interestingly, the *self-perception* literature in psychology (starting with Bem, 1967a,b) argues that humans do not have conscious access to their own preferences and uncover them in the same way as an external observer would, by updating them in response to their own observed choices. This view would be aligned with our first example, with the (Bayesian) observer and the agent again being different aspects of the same decision maker.

prior and posterior choice probabilities, and the updating rule used to transition from the former to the latter.

The problem we tackle goes beyond the characterization of choice data. In particular, we show that, even if a given set of choice frequencies can be described by alternative, formally equivalent models, the equivalence breaks down when updating is considered, because the structural assumptions will condition the posterior model. That is, two otherwise-equivalent models, confronted with the same new information, will result in different predictions for posterior choices. For example, consider random preference models (McFadden and Richter, 1990), which describe choices as the result of a probability distribution over deterministic preferences. For a finite set of alternatives, it is well-known that these models can alternatively be seen as *random utility models* (RUMs), which postulate a fixed utility function that is perturbed by an added noise term (Marschak, 1960, McFadden, 2001). For particular distributions of the noise, the latter encompass logit choice and the model of Luce (1959). Thus, the same dataset can be rationalized by a random preference model (a distribution over deterministic preferences) and a random utility model (an underlying utility function plus a specification of noise). The parameters of those models, though, are radically different, and updating in terms of these parameters, even on the basis of the same data, will in general result in different updated choice probabilities. This is why a dataset including histories can distinguish between otherwise-equivalent models of stochastic choice, and why their respective characterizations will differ.

After formulating the general framework, we proceed to characterize the most important cases in terms of first principles (axiomatic properties) which help clarify the differences between alternative approaches, and offer new insights on standard approaches to stochastic choice. On one hand, we consider the *logit-reinforcement model*, which relies on a given utility or value function which is updated according to some reinforcement process, and predicts stochastic choices through a logit choice function. We show that this model (which lies at the foundation of modern decision neuroscience) is characterized by a set of axioms on (posterior) choice probabilities. The first two key axioms state that updated choice probabilities for a chosen alternative do not depend on the identity of the alternative that was rejected, and that when an alternative is not chosen, its choice probabilities against third alternatives do not change. The third key axiom is a form of history independence: the odds of choosing one alternative over another, given a previous history of choices, depends on that history only through the most recent choice. Together with classical properties due to Luce and Suppes (1965), these axioms fully characterize the logit-reinforcement model.

On the other hand, we consider the *RU-Bayesian model*, which postulates a prior probability distribution over strict preferences, to be updated according to Bayes' rule in the face of new choice observations. It is well-known that multiple probability distributions might generate the same choice probabilities, creating an identification problem. We show that this problem is pervasive, in the sense that updating does not in general resolve the multiplicity. That is, there exist different probability distributions over preferences which induce the same choice frequencies even after updating, and do so even *for any possible choice observation.* Intuitively, however, this problem must be resolved after enough rounds of updating, and indeed we show that in the full dynamic case, a single axiom characterizes the RU-Bayesian model and solves the identification problem. This *Axiom of Bayesian Stochastic Preference* requires that the product of updated probabilities along a given history equals the sum of analogous products along all maximal histories (in a well-defined sense) which can extend the given history without contradicting it.

We also ask whether the models we consider can capture choice-induced preference. While some of the properties implied by this theory are satisfied both by the logit-reinforcement and the RU-Bayesian models, the key prediction arising from choice-induced preference change

is that the probability of choosing a previously-chosen alternative must increase even if the new choice is made against a new alternative. This prediction is indeed consistent with the logit-reinforcement model, but contradicts the RU-Bayesian model. Taken as a condition on primitives, it characterizes *positive* reinforcement within the class of logit-reinforcement models. This illustrates how our analysis clarifies which updating models can be used for which research questions.

The paper is organized as follows. Section 2 relates our contribution to the recent literature. Section 3 describes our general approach to stochastic choice models with updating and introduces the two benchmark classes: the logit-reinforcement model and the RU-Bayesian model. Section 4 presents the characterization of the logit-reinforcement model, and Section 5 does the same for the RU-Bayesian model. This section also discusses how our results allow to disentangle different models with updating, beyond the two benchmark classes. Section 6 discusses the implications for preference change. Section 7 concludes. We prove theorems and propositions in the paper, but defer the proof of corollaries and other straightforward calculations to the Appendix.

2. Related Literature

An extensive literature in both economics and psychology has concerned itself with the rationalizability of choice datasets. Specifically, the question asked is under which circumstances can a given set of hypothetical or actual choices be reproduced (rationalized) as the result of a specific theoretical model generating stochastic choices. Our work is related to a recent strand of the literature which considers how enriched datasets can be modeled and explained through models of stochastic choice. For instance, Caplin and Martin (2015) and Caplin and Dean (2015) consider state-dependent datasets, which specify choice frequencies as functions of observable states, and ask when those frequencies can be rationalized in terms of utility maximization with Bayesian updating based on imperfect signals. Caplin and Martin (2015) take the mapping from states to signals as given, while Caplin and Dean (2015) internalize it in terms of a rational inattention model. In contrast to our work, in those contributions updating is necessarily Bayesian and does not take previous choices as data, but rather signals on underlying states of the world. Yet, we share the basic motivation of rationalizing choice datasets including how the model is updated.

In Caplin and Martin (2015) and Caplin and Dean (2015) datasets are *extended* in the sense that the observer knows the states of the world and has access to choice frequencies conditional on those. In our case, the extension arises from the possibility to condition choice frequencies on previous choices. In this sense, our work is also related to other approaches considering datasets which add new dimensions to given choices. For instance, Alós-Ferrer, Fehr, and Netzer (2020) consider datasets where, in addition to choice frequencies, response times are also available (and hence must be rationalized by the model).

Our work is also related to the small but growing strand of the literature examining the axiomatic foundations of dynamic stochastic choice. Two important recent contributions are Fudenberg and Strzalecki (2015), which is related to our analysis of the logit-reinforcement model, and Frick et al. (2019), which is related to our analysis of the RU-Bayesian model. Fudenberg and Strzalecki (2015) consider an intertemporal choice framework where the decision maker chooses from a menu of actions yielding an outcome for today and another menu of actions for tomorrow. They provide an axiomatic characterization of a Discounted Adjusted Luce representation, which extends the classic logit model to a dynamic setting while allowing for aversion to larger choice sets. Choice in each period is stochastic due to random i.i.d. preference shocks. However, in contrast to our work, stochastic choice is not history-dependent, and

the utilities of outcomes are constant over time. As a consequence, Fudenberg and Strzalecki (2015) excludes the type of reinforcement dynamics that are at the center of our analysis for the dynamic logit choice model.

Frick et al. (2019) study dynamic random expected utility, extending the characterization of random expected utility in Gul and Pesendorfer (2006) to an intertemporal lottery framework. In their model, the decision maker's expected utility over lotteries (yielding an outcome for today and menu of lotteries for tomorrow) is state dependent, and uncertainty about the state is resolved over time. As such, their primitive is a dynamic history-dependent stochastic choice correspondence, and they provide axioms that characterize when such data can be rationalized by a model of state-dependent dynamic expected utility maximization. We also look at history-dependent stochastic choice, but both our primitives and our RU-Bayesian model differ substantially from theirs. In particular, a framework with lotteries over dynamic menu-choice problems is central to the analysis of expected utility maximization in Frick et al. (2019). In contrast, we consider a more abstract dynamic choice environment, with repeated binary choice problems for an arbitrary set of alternatives. The key axioms in Frick et al. (2019) are therefore either vacuous in our framework (contraction history independence), or cannot be translated into it (linear history independence and the axioms of Gul and Pesendorfer, 2006). Instead, we provide a simple characterization of Bayesian updating for arbitrary (strict) preferences on an abstract, finite set of alternatives, and show how the predictions of the model can be used to distinguish stochastic choice generated by the RU-Bayesian model from alternative models such as logit-reinforcement.

3. Stochastic Choice Models with Updating

In this section, we develop a binary-choice framework to study updating of stochastic choice models based on new, observed choices. We first review the standard framework for stochastic choice, and then extend the framework to incorporate updating. Then, we present the two benchmark classes (logit-reinforcement and RU-Bayesian models) and use them to show that different models initially rationalizing the same choice probabilities can make different predictions on the basis of the exact same observed choices.

3.1. *Stochastic Choice*

Let $A = \{a, b, ...\}$ be a finite set of *alternatives*, and let $\mathcal{A} = \{(a, b) \in A^2 : a \neq b\}$ be the set of all binary choices, i.e., ordered pairs of distinct alternatives. The ordered pair (a, b) is interpreted as the observation that, when confronted with the binary choice problem $\{a, b\}$, the decision maker chooses a. The observed, predicted, or hypothetical choice probabilities describing the decisions of an individual or a population are summarized by a *stochastic choice function*.

DEFINITION 1: A (binary) *stochastic choice function* (SCF) is a mapping $P : \mathcal{A} \to [0, 1]$ such that $P(a, b) = 1 - P(b, a)$ for all $(a, b) \in \mathcal{A}$.

The interpretation of the SCF is that $P(a, b)$ is the probability that the decision maker chooses alternative a from the choice set $\{a, b\}$, or, alternatively, the proportion of agents in a population who choose a when facing that binary choice set. Conversely, $P(b, a) = 1 - P(a, b)$ is the probability (or frequency) of the choice of b from the same choice set. A SCF can be taken to describe a data set, where $P(a, b)$ and $P(b, a)$ are the observed empirical frequencies of choices, or the predictions of a particular stochastic choice model.

While our work is ultimately about the rationalizability of SCFs, the question we ask is more demanding, namely when can choice probabilities *and their updated values given previously-observed choices* be rationalized by specific models. For our purposes, it is useful to formally introduce the concept of a stochastic choice model at this point, which corresponds simply to families of SCFs sharing a common structure.

DEFINITION 2: A *stochastic choice model* (SCM) is a family of SCFs indexed on a nonempty set Θ, that is, $\{P_\theta\}_{\theta\in\Theta}$. A given SCF P can be *rationalized* by the stochastic choice model if there exists $\theta_0 \in \Theta$ such that $P_{\theta_0} = P$. The value θ_0 is then referred to as the *parameter* of the SCM $\{P_\theta\}_{\theta\in\Theta}$ that *generates* the SCF P.

The following examples briefly review some prominent stochastic choice models, both as an illustration and for later reference.

EXAMPLE 1: A SCF P can be rationalized by the (binary) *Luce model* (Luce, 1959) if there exists a strictly positive (utility) function $v : A \mapsto \mathbb{R}_{++}$ such that

$$P(a,b) = \frac{v(a)}{v(a)+v(b)} \quad \text{for all } (a,b) \in \mathcal{A}. \tag{1}$$

The Luce model is a stochastic choice model with $\Theta_{\text{Luce}} = \{v \mid v : A \mapsto \mathbb{R}_{++}\}$, the set of all strictly positive utility vectors. By using the transformation $u(a) = \ln v(a)$, the equation can be rewritten as

$$P(a,b) = \frac{e^{u(a)}}{e^{u(a)}+e^{u(b)}} \quad \text{for all } (a,b) \in \mathcal{A}, \tag{2}$$

with $u : A \mapsto \mathbb{R}$ a (not necessarily positive) real-valued function, which can be interpreted as a utility function representing deterministic underlying preferences. This is the well-known (binomial) *logit model* (e.g., McFadden, 2001). Hence, the logit model is a SCM with $\Theta_{\text{Logit}} = \{u \mid u : A \mapsto \mathbb{R}\}$. By virtue of the logarithmic transformation, the Luce and the logit models are equivalent in the sense that a SCF can be rationalized by one model if and only if it can be rationalized by the other model.

EXAMPLE 2: A SCF P can be rationalized by the *Fechnerian model* (see e.g. Moffatt, 2015) if there exist a utility function $u : A \mapsto \mathbb{R}$ and a cumulative distribution function (cdf) $G : \mathbb{R} \mapsto [0,1]$ such that $P(a,b) = G(u(a) - u(b))$ for all $(a,b) \in \mathcal{A}$.

The Fechnerian model is a SCM with $\Theta_{\text{Fechner}} = \mathcal{G} \times \{u \mid u : A \mapsto \mathbb{R}\}$, where $\mathcal{G}$ is the set of cdfs. Of course, for each fixed cdf G, one obtains a particular submodel, which is a SCM in its own right, with $\Theta_G = \{u \mid u : A \mapsto \mathbb{R}\}$. A particularly well-known example, the *probit model*, is given when $G = \Phi$, where Φ is the cdf of a standard normal distribution. Taking $G(x) = 1/(1+e^{-x})$ instead generates the *logit model* of Example 1. That is, every SCF which is rationalizable by the Luce or logit model is also rationalizable by a (particular) Fechnerian model.

EXAMPLE 3: A SCF P can be rationalized by the *Random Utility Model (RUM)* if there exists a vector of utilities $u \in \mathcal{U}$ and a vector of jointly distributed random variables $(\varepsilon_a)_{a\in A}$ such that $P(a,b) = \text{Prob}\left(u(a)+\varepsilon_a \geq u(b)+\varepsilon_b\right) = \text{Prob}\left(u(a)-u(b) \geq \varepsilon_b - \varepsilon_a\right)$.

The RUM has gained considerable prominence in decision theory, discrete choice econometrics, and many branches of applied economics (Anderson, Thisse, and De Palma, 1992).[3] Of course, one can obtain particular submodels by fixing the distributions of noise, and, abusing notation, we could also refer to the resulting submodel as a (particular) RUM. For instance, if the random variables $\varepsilon_b - \varepsilon_a$ are identically distributed according to a cdf G, then $P(a,b) = G(u(a) - u(b))$ and the resulting submodel is a Fechnerian model. For instance, if the ε_a are normally distributed, so are the $\varepsilon_b - \varepsilon_a$, and one obtains the probit model. If the ε_a follow a double exponential distribution, then the $\varepsilon_b - \varepsilon_a$ follow a logistic distribution and one obtains the logit model.

The RUM can also be expressed in a different way that, for our purposes, gives rise to a particularly relevant index set. Let $\mathcal{R}$ denote the set of strict preference orderings on $\mathcal{A}$, with typical element $\succ \in \mathcal{R}$. For each $\succ \in \mathcal{R}$, define the (degenerate) SCFs

$$P_{\succ}(a,b) = \begin{cases} 1 & \text{if } a \succ b \\ 0 & \text{if } b \succ a. \end{cases}$$

Let $\Theta = \Delta(\mathcal{R})$ denote the set of probability distributions on $\mathcal{R}$. For each distribution $\pi \in \Delta(\mathcal{R})$, define $P_\pi(a,b) = \sum_{\succ \in \mathcal{R}} \pi(\succ) P_{\succ}(a,b)$. Following the terminology of McFadden and Richter (1990), we refer to this approach as the *random preference model*. Block and Marschak (1960) showed that the RUM and the random preference model are equivalent, in the sense that choice probabilities for a finite set of alternatives can be rationalized by the RUM if and only they can be rationalized by the random preference model. In that sense, the random preference model is just a different way of conceptualizing the RUM.

3.2. *Updating Choice Probabilities*

We now turn to the description of stochastic choice functions incorporating updated choice probabilities on the basis of previous choices. Choice probabilities correspond to a family of SCFs, one describing the prior probabilities of choices, and others describing the updated (conditional) probabilities after any given history of choices has been observed. A subtlety, however, is that the latter should exclude counterfactuals, that is, they only describe probabilities after the observations of choices that actually had a positive probability given previously observed choices.

For notational convenience, let $\mathcal{A}^* = \mathcal{A} \cup \{\emptyset\}$ be the set of observable binary choices enriched with a distinguished element, denoted $\emptyset$, which indicates that no choice has been observed. Hence, $P(\cdot|\emptyset)$ below will refer to the prior. Now suppose that we are confronted with several observed choices. A *history of length* $n \geq 1$ is an ordered collection (i.e., a vector) of n observations from $\mathcal{A}$, $h = (s_1, \ldots, s_n) \in \mathcal{A}^n$, with the interpretation that s_1 is the most-recent choice observed ("one period ago"), s_2 is the choice observed before s_1, etc. A *history* is an ordered collection of choice observations, of arbitrary but finite length. The set of all histories is $\mathcal{H} = \bigcup_{n=0}^{\infty} \mathcal{A}^n$, where $\mathcal{A}^0 = \{\emptyset\}$ by convention. Given a history $h \in \mathcal{H}$, denote its *length* by $\ell(h)$, i.e. $\ell(h) = n$ if and only if $h \in \mathcal{A}^n$. For convenience, given $s \in \mathcal{A}$ and $h = (s_1, \ldots, s_N) \in \mathcal{H}$, denote by $s \circ h$ the history formed by following up history h with the choice s, i.e. $s \circ h = (s, s_1, \ldots, s_n)$.

[3] Standard terminology speaks of "RUMs," because every utility function and distribution of noise is a different model in an econometric sense. In our terms, the entire family of RUMs is a SCM, which we refer to as *the* random utility model, on the same footing as, say, the Luce model viewed as a SCM.

Updated choice probabilities must be specified for any history which can actually be observed according to the previous probabilities, in an iterative fashion. This is captured by the following definition.

DEFINITION 3: A *dynamic stochastic choice function* (DSCF) is a pair $(\bar{\mathcal{H}}, P)$ consisting of a set of *admissible histories*, $\bar{\mathcal{H}} \subseteq \mathcal{H}$, and a mapping $P : \mathcal{A} \times \bar{\mathcal{H}} \mapsto [0,1]$ such that
(i) $\emptyset \in \bar{\mathcal{H}}$,
(ii) for each $s \in \mathcal{A}$ and $h \in \mathcal{H}$, $s \circ h \in \bar{\mathcal{H}}$ if and only if $h \in \bar{\mathcal{H}}$ and $P(s|h) > 0$, and
(iii) for each $h \in \bar{\mathcal{H}}$, the induced mapping $P(\cdot|h) : \mathcal{A} \mapsto [0,1]$ is a SCF.

The interpretation of the DSCF is as follows. First, $P(\cdot|\emptyset)$ describes the prior choice probabilities. Then, $P(a,b|h)$ is the posterior probability that the decision maker chooses alternative a from the choice set $\{a,b\}$ given a previous *collection* of choices as described in the history h. Again, $P(\cdot|h)$ might include zero values, but the history h must be feasible in the sense that each choice along h had positive probability given the preceding choices.

The question we ask is when can a DSCF be rationalized by a specific model of stochastic choice which allows for updating. The following definition describes such models.

DEFINITION 4: A *stochastic choice model with updating* (SCM-U) is defined by a tuple $(\Theta, \{P_\theta\}_{\theta\in\Theta}, f)$ where Θ is a nonempty index set and $f : \Theta_f \mapsto \Theta$ is a *choice-updating function*, such that
(i) P_θ is a SCF for each $\theta \in \Theta$,
(ii) $\Theta_f = \{(\theta, s) \in \Theta \times \mathcal{A}^* \mid s = \emptyset \text{ or } P_\theta(s) > 0\}$, and
(iii) $f(\theta, \emptyset) = \theta$ for all $\theta \in \Theta_f$.

The interpretation of a SCM-U is that the decision maker has (or is believed to have) some prior parameter $\theta_0 \in \Theta$ before making any choices, but after making an initial choice of c from $\{c,d\}$, the parameter is updated to $\theta' = f(\theta_0, (c,d))$. Hence, her new (actual or predicted) choice probabilities are given by the new SCF $P_{\theta'}$, and she will choose a from $\{a,b\}$ with probability $P_{\theta'}(a,b)$, instead of the original $P_{\theta_0}(a,b)$.

Updating then proceeds iteratively when additional choices are observed. To describe this iterative process, consider the following notation. Given a choice-updating function f as in Definition 4, write $f(\theta,h) = f(\theta,s)$ if $h = (s) \in \mathcal{A}$, and, iteratively, for each $n \geq 2$, $f(\theta,h) = f(f(\theta,(s_2,\ldots,s_n)),s_1)$ for $h = (s_1,s_2,\ldots,s_n) \in \mathcal{A}^n$. Thus, the function f yields an updated parameter $f(\theta,h)$ following any history of choices h and the updated choice probabilities (or frequencies) after history h are described by the SCF $P_{f(\theta,h)}$.

REMARK 1: By the above iterative process, any $\theta_0 \in \Theta$ generates a DSCF with a set of admissible histories $\bar{\mathcal{H}}$ (defined iteratively) and history-dependent SCFs defined by $P(.|h) \equiv P_{f(\theta,h)}(.)$ for all $h \in \bar{\mathcal{H}}$. Hence, we say that θ_0 is the *prior parameter* of the SCM-U $(\Theta, \{P_\theta\}_{\theta\in\Theta}, f)$ that *generates* the DSCF $(\bar{\mathcal{H}}, P)$. Since every prior parameter generates a DSCF, a SCM-U can be interpreted simply as a collection of DSCFs, providing a dynamic analog of SCMs as a collection of SCFs in Definition 2.

The following definition formalizes when an SCM-U rationalizes a given DSCF.

DEFINITION 5: A DSCF $(\bar{\mathcal{H}}, P)$ can be rationalized by a SCM-U $(\Theta, \{P_\theta\}_{\theta\in\Theta}, f)$ if there exists $\theta_0 \in \Theta$ such the SCM-U with the prior parameter θ_0 generates the DSCF, i.e., $P(a,b|h) = P_{f(\theta_0,h)}(a,b)$ for all $(a,b) \in \mathcal{A}$ and $h \in \bar{\mathcal{H}}$.

Whenever a DSCF can be rationalized by a SCM-U, the model provides not only the structure linking underlying parameters (e.g., utilities or beliefs) to stochastic choices, but also the structure of the updating process. The following proposition shows that every DSCF admits such a rationalization, showing that our definition of SCM-Us is general enough to encompass all possible DSCFs in an abstract binary-choice framework.

PROPOSITION 1: *Every DSCF can be rationalized by some SCM-U.*

PROOF: Let $(\bar{\mathcal{H}}, P)$ be a DSCF. To construct a SCM-U that rationalizes $(\bar{\mathcal{H}}, P)$, let $\Theta = \bar{\mathcal{H}}$, so that each $\theta \in \Theta$ corresponds to an admissible history of the DSCF. To define the collection of SCFs $\{P_\theta\}_{\theta\in\Theta}$, consider $\theta = h$ and let $P_\theta(.) = P(.|h)$, where the right-hand-side is the stochastic mapping of the given DSCF. Finally, define the updating function $f : \Theta_f \to \Theta$ as follows. First, for $\theta = h$ and $s \in \mathcal{A}^*$, let $(\theta, s) \in \Theta_f$ if and only if $s \circ h \in \bar{\mathcal{H}}$, which defines the nonempty set Θ_f. Finally, for $(\theta, s) = (h, s) \in \Theta_f$, let $f(\theta, s) = s \circ h$, which is in Θ by definition. It is then easily verified that the SCM-U $(\Theta, \{P_\theta\}_{\theta\in\Theta}, f)$ rationalizes the DSCF $(\bar{\mathcal{H}}, P)$ in terms of the prior parameter $\theta_0 = \emptyset$. *Q.E.D.*

Together with the observations in Remark 1, Proposition 1 can be interpreted as an abstract representation result: SCM-Us provide a parsimonious, structural representation for DSCFs. Our main results identify the axiomatic properties of those DSCFs that can be rationalized by two specific SCM-Us that are of particular interest given the prominence of the logit and random utility models in the literature on stochastic choice.

3.3. *Logit Models with Reinforcement Learning*

We first consider a stochastic choice model with updating that is especially important in view of recent developments in neuroeconomics. Extensive evidence from neuroscience (see, e.g., Holroyd and Coles, 2002, Schultz, 2013, Daw and Tobler, 2014) shows that the dopaminergic system in the brain reflects value learning through elementary reinforcement processes. Indeed, reinforcement is generally viewed as the most basic learning process for human beings, and accordingly has received a great deal of attention in economics, psychology, and other fields as computer science (Sutton and Barto, 1998). The classical formal models of reinforcement learning are those of Bush and Mosteller (1951, 1955), first introduced to economics by Cross (1973, 1983). In those models, the probability of choosing an action is adjusted by a quantity proportional to the reward resulting from that action. For instance, when all possible rewards are positive, the model can be seen to capture habit formation. An analytically inconvenient characteristic of those models, however, is that the probabilities of actions not chosen also need to be adjusted to preserve the total sum. However, this formal difficulty is bypassed if reinforcement acts on some abstract propensities which are converted into probabilities by an appropriate function, as e.g. any Luce-like functional form. Hence, when considering Luce or logit models, it is natural to consider an adjustment of choice probabilities through reinforcement of the values $u(x)$.

Formally, denote by $\mathcal{U} = \{u \mid u : A \mapsto \mathbb{R}\}$ the set of all real-valued (utility) functions on A, and fix a strictly increasing transformation $\omega : \mathbb{R} \mapsto \mathbb{R}_{++}$. Let $\Theta = \mathcal{U}$ and, for each $u \in \mathcal{U}$, define $P^\omega_u(a, b) = \frac{\omega(u(a))}{\omega(u(a))+\omega(u(b))}$. Letting $v = \omega \circ u$, this formula describes a different parameterization of the Luce model. If $\omega(x) = e^x$, it corresponds to the logit model. As observed in Example 1, every SCF which can be rationalized by the Luce model can also be rationalized by the logit model.

Now consider a function $R : A \to \mathbb{R}$, interpreted as a reinforcement function, and define the choice-updating function $f_R : \Theta \times \mathcal{A}^* \to \Theta$ by $f_R(u, s) = u_s$ where, for all $a \in A$,

$$u_s(a) = \begin{cases} u(a) & \text{if } s = \emptyset \text{ or } s = (c, d) \text{ with } c \neq a \\ u(a) + R(a) & \text{if } s = (c, d) \text{ with } c = a. \end{cases}$$

For any $\omega : \mathbb{R} \mapsto \mathbb{R}_{++}$ and function R as above, the tuple $(\mathcal{U}, \{P_u^\omega\}_{u \in \mathcal{U}}, f_R)$ describes a stochastic choice model with updating which we refer to as the *Luce-reinforcement model* with reinforcement function R. When $\omega(x) = e^x$, we refer to it as the *logit-reinforcement model with reinforcement function* R.

A case of particular interest is *positive reinforcement*, where $R(a) > 0$ for all $a \in A$, which is natural if none of the alternatives is aversive, or in models of habit formation. In general, however, reinforcement could be positive or negative. This does not necessarily mean that the options themselves deliver negative or unpleasurable experiences, as models with negative reinforcers arise naturally in the neuroeconomics literature.

EXAMPLE 4: A standard model in neuroeconomics, based on extensive neural evidence, posits that reinforcers correspond to *Reward Prediction Errors* (e.g., Schultz, Dayan, and Montague, 1997, Daw and Tobler, 2014). Let u represent the current decision value, which is used to make decisions. The brain employs it as a *predicted* reward (implemented by a brain network relying on the neurotransmitter dopamine; Platt and Glimcher, 1999, Schultz, 2010, 2013). However, an actual reward $r(a)$ is experienced when a is chosen from $\{a, b\}$. The reward prediction error is $RPE(a) = r(a) - u(a)$, i.e. the difference between the experienced and the predicted values (for an axiomatic characterization of such functions, see Caplin and Dean, 2008 and Caplin et al., 2010). The decision value is then updated through $u_s(a) = u(a) + \alpha \cdot RPE(a) = (1 - \alpha)u(a) + \alpha r(a)$, where $\alpha \in (0, 1)$ is the speed of adjustment. This model is a particular case of reinforcement with $R(a) = \alpha RPE(a)$. Even if $r(a) > 0$ for all $a \in A$, the reward prediction error can be positive or negative.

In Section 4, we characterize the DSCFs which can be rationalized by a logit-reinforcement model, as well the special case of positive reinforcement.

3.4. *The Random Utility Model with Bayesian Updating*

Our second main model builds on the RUM. As discussed in Example 3, the RUM can equivalently be formulated in terms of the random preference model. The difference, however, is consequential for our approach, for the random preference view defines a SCM $\{P_\pi\}_{\pi \in \Delta(\mathcal{R})}$, with P_π as given in the Example 3; that is, it indexes the different random preferences through the distributions over deterministic preferences. This formulation facilitates a natural interpretation. Imagine an external observer is unsure about the preferences of a decision maker, and learns about them by observing the decision maker's choices. Then the distribution π can be seen as the prior distribution capturing the observer's initial beliefs over the decision maker's preferences. Learning then reduces to defining an appropriate mapping over the space $\Theta = \Delta(\mathcal{R})$. Formally, the model also encompasses the idea of *self-perception* from psychology, which originates with Bem (1967a,b) (see Section 6 below). Imagine a decision maker is unsure about his or her *own* preferences, and learns their preferences by observing their own choices, in very much the same way as he or she would learn the preferences of a different agent. Then the distribution π is the prior distribution capturing the agent's self-model.

Formally, consider the random preference model $\{P_\pi\}_{\pi\in\Delta(\mathcal{R})}$ as defined in Example 3. This model is a SCM with $\Theta = \Delta(\mathcal{R})$. Rational learning in this setting considers an observed choice $(a,b) \in \mathcal{A}$ as new evidence allowing for a revision of the prior π by Bayesian updating. Hence, we define the Bayesian-updating function $f_B : \Delta(\mathcal{R}) \times \mathcal{A}^* \to \Delta(\mathcal{R})$ by

$$f_B(\pi, s) = \begin{cases} \pi & \text{if } s = \emptyset \\ \pi_s & \text{if } s = (a,b) \in \mathcal{A}, \end{cases} \tag{3}$$

where, for each $(a,b) \in \mathcal{A}$ with $P_\pi(a,b) > 0$,

$$\pi_{a,b}(\succ) = \begin{cases} \frac{\pi(\succ)}{P_\pi(a,b)} & \text{if } a \succ b \\ 0 & \text{if } b \succ a. \end{cases} \tag{4}$$

The tuple $(\Delta(\mathcal{R}), \{P_\succ\}_{\succ\in\mathcal{R}}, f_B)$ is a stochastic choice model with updating, which we refer to as the *RU-Bayesian model.* Note that a choice-updating function as given in Definition 4 is only specified for observations which can actually be observed in the SCM-U. Accordingly, the Bayesian updating function f_B is only defined for pairs $(\pi, (a,b)) \in \Delta(\mathcal{R}) \times \mathcal{A}$ such that $P_\pi(a,b) > 0$, since it is not specified for counterfactuals, i.e. if an observation (a,b) is not possible under π.

In Section 5, we characterize the DSCFs which can be rationalized by the RU-Bayesian model.

3.5. *Updating Under Different Models*

A given collection of choice probabilities can often be rationalized according to different stochastic choice models. For example, suppose choice probabilities are derived from a utility function u and a logit choice function. This corresponds to a random utility model for a specific distribution of errors (e.g., McFadden, 2001). Hence, by the equivalence result of Block and Marschak (1960), there exists a probability distribution over strict preferences which, if interpreted as a random preference model, generates exactly the original choice frequencies. Thus, the modeler can equivalently adopt one view or the other, as long as he or she adopts a static view and does not consider further updating.

To fix ideas, suppose we have a set of three alternatives $A = \{a,b,c\}$ and we are given choice probabilities $P(a,b) = 2/3$, $P(a,c) = 4/5$, and $P(b,c) = 2/3$. These probabilities can be rationalized by a utility u such that $u(a) = \ln 4$, $u(b) = \ln 2$, and $u(c) = 0$, and the logit choice function in (2). However, a straightforward computation shows that the same choice probabilities can be rationalized as resulting from a distribution π over strict preferences given by the following table:

Preference	$a \succ b \succ c$	$a \succ c \succ b$	$b \succ a \succ c$	$b \succ c \succ a$	$c \succ a \succ b$	$c \succ b \succ a$
π	1/3	1/3	2/15	1/5	0	0

The modeler does not need to take a stance on which model to favor, as they are both equivalent in terms of stochastic choices. Suppose, however, that a new choice is observed, say (a,b). If the modeler is a Bayesian observer employing the random preference model, he or she will update the prior by applying Bayes' rule, hence relying on the RU-Bayesian model from Section 3.4. Trivially, this results in a posterior distribution placing probability $1/2$ on each of the two preferences $a \succ b \succ c$ and $a \succ c \succ b$. Hence, the updated choice probabilities

are $P_B(a,b|a,b)=1$, $P_B(a,c|a,b)=1$, and $P_B(b,c|a,b)=1/2$. Alternatively, the modeler could also decide on the logit model and adopt a reinforcement approach based on some reinforcement function R, hence relying on the logit-reinforcement model from Section 3.3. Let $r=e^{R(a)}$. Utility is then updated to $u'(a)=\ln 4+\ln r=\ln 4r$, $u'(b)=\ln 2$, and $u'(c)=0$, and hence the updated choice probabilities are given by

$$P_R(a,b|a,b)=\frac{2}{2+(1/r)} \qquad P_R(a,c|a,b)=\frac{4}{4+(1/r)} \qquad P_R(b,c|a,b)=\frac{2}{3}.$$

The two models, therefore, produce different updated choice probabilities when confronted with the same new evidence. This results from the requirement that updating occurs *within the model*, which creates different modes of updating. Since the parameter (in our terms, θ) in a logit SCM is a utility function, it is that function which is updated, and this is the essence of reinforcement. In contrast, the parameter in a random preference model is a distribution over strict preferences, and the natural way to update is using Bayes' rule.

Given the formal framework developed above, this observation becomes straightforward. However, it is worth noticing that the differences are of a fundamental nature. For instance, if one considers reinforcement models with $r\to\infty$, in the limit $P_R(a,b|a,b)\to 1=P_B(a,b|a,b)$ and $P_R(a,c|a,b)\to 1=P_B(a,c|a,b)$, i.e., the logit-reinforcement predictions approach the Bayesian ones for these binary choice pairs. However, for any value of r, $P_R(b,c|a,b)=2/3\neq 1/2=P_B(b,c|a,b)$, and so the posterior SCFs $P_R(.|a,b)$ and $P_B(.|a,b)$ are distinct even in the limit. The axiomatic characterizations in the next two sections will clarify the structural differences among the models.

4. The Logit-Reinforcement Model

In view of the prominence of the logit choice model and the empirical relevance of reinforcement learning, we proceed to characterize those DSCFs that can be rationalized by a logit-reinforcement model. We first recall the characterization of SCFs (without updating) that can be rationalized by a logit model. We then provide axioms that characterize a logit model with reinforcement learning, with an additional result identifying the case of positive reinforcement. Finally, we consider the generalization to Fechnerian models.

4.1. *Logit Stochastic Choice Functions*

Luce (1959, 1977) formulated stochastic choice models for choices from arbitrary sets (i.e., the primitives are $P(b|B)$ for $b\in B\subseteq A$) and characterized choice rules of the form (1) through two axioms. The first is *positivity* (all choice probabilities are strictly positive). The second is the celebrated *Luce's choice axiom*, which states that the probability that an option is chosen from a set does not change if an intermediate subset containing that alternative is chosen first, and then choice is restricted to that subset. One important implication of this property, often called *independence of irrelevant alternatives*, is that the ratio of choice probabilities between two alternatives a,b does not depend on which choice set is considered, provided it contains both a and b. Since we focus on binary choice, the characterization in Luce (1959) does not apply, because the choice axiom is void in this case: given $a,b\in A$, there exists only one possible choice set containing both, namely $\{a,b\}$. However, Luce's characterization can be adapted to a binary choice framework by replacing the choice axiom with *Luce's Product Rule*. The two axioms are:

AXIOM—POS: *For all* $(a,b)\in\mathcal{A}$, $P(a,b)>0$.

AXIOM—LPR: *For all distinct* $a, b, c \in A$, $P(a,b)P(b,c)P(c,a) = P(a,c)P(c,b)P(b,a)$.

The product rule has a natural interpretation in terms of the probabilities of observing cycles in choices (see, e.g., Baldassi et al., 2020). Given that choice probabilities are independent, $P(a,b)P(b,c)P(c,a)$ can be interpreted as the probability of observing the intransitive cycle $a \to b \to c \to a$, while $P(a,c)P(c,b)P(b,a)$ can be interpreted as the probability of observing the cycle $a \to c \to b \to a$. The LPR then asserts that the probability of observing these choice cycles should be the same, and so violations of transitivity are not systematic, but rather due to pure noise.

The following result from Luce and Suppes (1965, p. 341) shows that POS and LPR characterize stochastic choice functions (without updating) that can be rationalized by a logit choice model. Here, we adapt the result to our notation, and provide a proof in Appendix A only for completeness.

LEMMA 1: *A SCF* P *can be rationalized by the Luce model (or by the logit model) if and only if* P *satisfies POS and LPR.*

For later reference, we also observe that the logit model is always identified up to an additive utility constant (again, this is well-known and we provide a proof in Appendix A only for completeness).

LEMMA 2: *Suppose a SCF can be rationalized by the logit model with parameter (utility)* u*, and also with parameter* u'*. Then, there exists a constant* K *such that* $u'(a) = u(a) + K$ *for all* $a \in A$*.*

As a result, one could define the logit model on the quotient set of the parameter space (utility functions) given by the equivalence relation where two utility functions are equivalent if one is a constant shift of the other. With this approach, the logit model is always fully identified in the sense that a rationalizable SCF would always correspond to one and only one value of the parameters in this quotient set (an equivalence class).

4.2. *Axioms for the Logit-Reinforcement Model*

We now state the axiomatic properties of a dynamic stochastic choice function $(\bar{\mathcal{H}}, P)$ that can be rationalized as logit (or Luce) models with (positive) reinforcement. In view of Lemma 1, the first two axioms state that updating occurs within the class of logit models; that is, the updated choice probabilities given any prior choices fulfill POS and LPR.

AXIOM—U-POS: *For all* $h \in \bar{\mathcal{H}}$, $P(\cdot|h)$ *satisfies POS.*

AXIOM—U-LPR: *For all* $h \in \bar{\mathcal{H}}$, $P(\cdot|h)$ *satisfies LPR.*

The following axioms are new and concern how choice probabilities are updated. The first, *stability of discarded alternatives* (SDA), states that the choice probabilities for a discarded alternative against third alternatives do not change. Note that this does not preclude that the choice probability of a discarded alternative changes if the same binary choice problem is encountered again.

AXIOM—SDA: *For all distinct* $a, b, c \in A$ *and all* $h \in \bar{\mathcal{H}}$, $P(b,c|(a,b) \circ h) = P(b,c|h)$.

The second new axiom, *independence of discarded alternatives* (IDA), states that the updated choice probabilities for a chosen alternative do not depend on the identity of the alternative which was not chosen.

AXIOM—IDA: *For all* $a,b,c,d \in A$ *such that* $a \notin \{b,c,d\}$ *and all* $h \in \bar{\mathcal{H}}$, $P(a,b|(a,c)\circ h) = P(a,b|(a,d)\circ h)$.

The third, *history independence* (HI), states that the percentual change in the odds of choosing an alternative a over another alternative, b, given a previous history h and given that a has just been observed to be chosen over an alternative c, does not depend on the previous history h.[4]

AXIOM—HI: *For all* $a,b,c,d \in A$ *such that* $a \neq b,c,d$ *and all* $h,h' \in \bar{\mathcal{H}}$,

$$\frac{\dfrac{P(a,b|(a,c)\circ h)}{P(b,a|(a,c)\circ h)}}{\dfrac{P(a,b|h)}{P(b,a|h)}} = \frac{\dfrac{P(a,b|(a,c)\circ h')}{P(b,a|(a,c)\circ h')}}{\dfrac{P(a,b|h')}{P(b,a|h')}} \tag{5}$$

whenever all involved probabilities are strictly positive.

The last axiom, *boosting of chosen alternatives* (BCA) states that the probability of an alternative to be chosen rises when the exact same choice is presented again. This axiom will help us distinguish between models with and without positive reinforcement.

AXIOM—BCA: *For all* $s \in \mathcal{A}$ *and all* $h \in \bar{\mathcal{H}}$ *with* $0 < P(s|h) < 1$, $P(s|s\circ h) > P(s|h)$.

Before providing the characterization results, we observe that, in the RU-Bayesian model, it is always the case that $P(s|s\circ h) = 1$ for $s\circ h \in \bar{\mathcal{H}}$ and so axiom BCA is (trivially) satisfied (see Section 5.4). However, all of the other new axioms (SDA, IDA, and HI) can, in general, be violated by DSCFs that can be rationalized by the RU-Bayesian model. We show this with a single example.

EXAMPLE 5: Suppose there are four alternatives $\mathcal{A} = \{a,b,c,d\}$ and consider the following prior probability over strict preferences:

Preference	$d \succ a \succ b \succ c$	$d \succ b \succ a \succ c$	$c \succ a \succ b \succ d$
π	1/3	1/3	1/3

Consider the RU-Bayesian model with the prior parameter π described above. To see that the DSCF generated by this model does not satisfy SDA or IDA, let $h = \emptyset$. Then, $P(b,c|a,b) = 1/2 \neq 2/4 = P(b,c|\emptyset)$, violating axiom SDA. Also, $P(a,d|a,b) = 1/2 \neq 0 = P(a,d|a,c)$, violating IDA. To see that HI fails, note that, with $h = \emptyset$, $P(a,b) = 2/3$ and $P(a,b|a,c) = 1/2$, thus the odds-quotient on the left-hand-side of Eq. (5) is $1/2$. However, if we let $h' = (d,c)$, then $P(a,b|h') = 1/2$ and $P(a,b|(a,c)\circ h') = 1/2$, and thus the quotient on the right-hand-side of Eq. (5) is equal to 1, violating HI.

[4] Axioms IDA and HI could be easily combined into one by stating independence from both the previous history h *and* the alternative c in HI's statement. We keep the axioms separate to make their respective role in our characterization more transparent.

4.3. *Characterization of Logit-Reinforcement Models*

We are now ready to state and prove our first main result (independence of the axioms is shown in Appendix B).

THEOREM 1: *A dynamic stochastic choice function $(\bar{\mathcal{H}}, P)$ can be rationalized by a logit model with reinforcement learning if and only if it satisfies U-POS, U-LPR, SDA, IDA, and HI. In particular, $\bar{\mathcal{H}} = \mathcal{H}$.*

PROOF: The necessity of the axioms is straightforward. To see the sufficiency, let $(\bar{\mathcal{H}}, P)$ be a DSCF that satisfies the five stated axioms. By U-POS, it follows that $\bar{\mathcal{H}} = \mathcal{H}$. We proceed in two steps: we first fix a history and consider the SCF after one additional choice observation, and then establish a connection between the SCFs given different histories.

Step 1. [One-step updating.] For a fixed history $h \in \mathcal{H}$, consider the choice probabilities given by $P^h(s|s') = P(s|s' \circ h)$. By Lemma 1, U-POS, U-LPR imply that, for every $s \in \mathcal{A}^*$, $P^h(\cdot|s)$ can be rationalized by a logit model: for each $s \in \mathcal{A}^*$, there exists $u_s \in \mathcal{U}$, such that

$$P^h(a,b|s) = \frac{e^{u_s^h(a)}}{e^{u_s^h(a)} + e^{u_s^h(b)}} \tag{6}$$

for all $(a,b) \in \mathcal{A}$. Note that u_s^h is not necessarily unique, but is unique up to the addition of a constant (Lemma 2), which does not impact our arguments. For notational simplicity, denote $u^h = u_\emptyset^h$ and $u_{a,b}^h = u_{(a,b)}^h$ for each $(a,b) \in \mathcal{A}^*$.

For any $a,b,c \in \mathcal{A}$ with $a \neq b,c$, define the *spreading* of (a,b) given c by

$$S(a,b,c) = \left(u_{a,b}^h(a) - u_{a,b}^h(c)\right) - \left(u(a) - u(c)\right).$$

We now observe some properties of this spreading function.

Step 1a. [$S(a,b,c)$ is independent of c.] By SDA, for all $a,b,c' \in A$ pairwise different, $P^h(b,c'|a,b) = P^h(b,c'|\emptyset)$. Hence, by (6), $u_{a,b}^h(c') - u(c') = u_{a,b}^h(b) - u^h(b)$ for all $c' \neq a,b$. Define $K(a,b) = u_{a,b}^h(b) - u^h(b)$. It follows that $u_{a,b}^h(c') - u^h(c') = K(a,b)$ for any $c' \neq a,b$, and in particular this quantity is independent of c'. Now note that

$$S(a,b,c) = \left[u_{a,b}^h(a) - u^h(a)\right] - \left[u_{a,b}^h(c) - u^h(c)\right] = \left[u_{a,b}^h(a) - u^h(a)\right] - K(a,b)$$

and is therefore independent of c for all $c \neq a,b$, as claimed. Moreover, if $c = b$,

$$S(a,b,b) = \left[u_{a,b}^h(a) - u^h(a)\right] - \left[u_{a,b}^h(b) - u^h(b)\right] = \left[u_{a,b}^h(a) - u^h(a)\right] - K(a,b).$$

Step 1b. [$S(a,b,c)$ is independent of b.] By IDA, for all $a,b,c,d \in A$ such that $a \neq b,c,d$, $P^h(a,d|a,b) = P^h(a,d|a,c)$. Hence, by (6) $u_{a,b}^h(d) - u_{a,b}^h(a) = u_{a,c}^h(d) - u_{a,c}^h(a)$, implying that $u_{a,b}^h(a) - u_{a,b}^h(d)$ is independent of b as long as $a \notin \{b,d\}$, i.e., $u_{a,b}^h(a) - u_{a,b}^h(d) = T(a,d)$ for some $T(a,d)$ independent of b. Note that $S(a,b,c) = T(a,c) - [u^h(a) - u^h(c)]$, and is therefore independent of b, as claimed.

Step 1c. [$S(a,b,c)$ is independent of (b,c).] Let $b,c,b',c' \in A \setminus \{a\}$. We have to show that $S(a,b,c) = S(a,b',c')$. By Step 1a, $S(a,b,c) = S(a,b,c')$. By Step 1b, $S(a,b,c') = S(a,b',c')$. Hence, the claim follows.

In view of Steps 1a–1c, for $a \in A$, we can define $R^h(a) = S(a,b,c)$ for any $\{b,c\} \not\ni a$ and we obtain that $R^h : A \to \mathbb{R}$ is well-defined. Further, for any $(a,b) \in \mathcal{A}$,

$$\begin{aligned} u_{a,b}^h(a) &= u^h(a) + \left[u_{a,b}^h(a) - u^h(a)\right] - K(a,b) + K(a,b) \\ &= u^h(a) + S(a,b,b) + K(a,b) = u^h(a) + R^h(a) + K(a,b). \end{aligned}$$

and $u^h_{a,b}(b) = u^h(b) + K(a,b)$. Further, for any $c \notin \{a,b\}$, by the argument in Step 1, $u^h_{a,b}(c) - u^h(c) = K(a,b)$, hence $u^h_{a,b}(c) = u^h(c) + K(a,b)$.

We now construct a choice-updating function by

$$f(u^h,(a,b))(c) = \begin{cases} u^h(a) + R^h(a) & \text{if } c = a \\ u^h(c) & \text{if } c \neq a. \end{cases}$$

It remains to show that, for all $(a,b),(c,d) \in \mathcal{A}$, $P^h(c,d|a,b) = P_{f(u^h,(a,b))}(c,d)$. To see this, note that

$$P^h(c,d|a,b) = \frac{1}{1+e^{u^h_{a,b}(d)-u^h_{a,b}(c)}} = \frac{1}{1+e^{f(u^h,(a,b))(d)-f(u^h,(a,b))(c)}} = P_{f(u^h,(a,b))}(c,d)$$

Step 2: [History independence of the reinforcement.] By Step 1, for each $h \in \mathcal{H}$ there exists a utility function u^h and a reinforcement function R^h such that, for all $(a,b) \in \mathcal{A}$,

$$P(a,b|h) = \frac{e^{u^h(a)}}{e^{u^h(a)} + e^{u^h(b)}} \tag{7}$$

and, for any $h^* = (a',b') \circ h$ with $(a',b') \in \mathcal{A}$, and for all $(a,b) \in \mathcal{A}$,

$$P(a,b|h^*) = \frac{e^{u^h_{h^*}(a)}}{e^{u^h_{h^*}(a)} + e^{u^h_{h^*}(b)}} \tag{8}$$

where

$$u^h_{h^*}(c) = \begin{cases} u^h(c) & \text{if } s = \emptyset \text{ or } s = (c,d) \text{ with } c \neq a \\ u^h(a') + R^h(a') & \text{if } s = (c,d) \text{ with } c = a'. \end{cases}$$

This identifies one utility function and one reinforcement function per history. We now show that the reinforcement functions are independent of history, $R^h = R^{h'}$ for all $h,h' \in \mathcal{H}$. To see this, note that, by (7) and (8), for $a \neq b,c$,

$$\frac{\frac{P(a,b|(a,c)\circ h)}{P(b,a|(a,c)\circ h)}}{\frac{P(a,b|h)}{P(b,a|h)}} = \frac{\frac{e^{u^h(a)+R^h(a)}}{e^{u^h(b)}}}{\frac{e^{u^h(a)}}{e^{u^h(b)}}} = \frac{e^{u^h(a)+R^h(a)}}{e^{u^h(a)}} = e^{R^h(a)}$$

and hence, by HI, $R^h(a) = R^{h'}(a)$ for all $h,h' \in \mathcal{H}$.

We have obtained a family of utility functions, $\{u^h \mid h \in \mathcal{H}^*\}$ and a history-independent reinforcement function, denoted now simply by R, such that (7) and (8) hold. To show that the DSCF can be rationalized by a logit-reinforcement model, it remains to be proven that we can assume the utility functions u^h to be derived from previous ones by using the reinforcement function R.

To show this, consider the utility function $u = u^\emptyset$. By (7), for all $(a,b) \in \mathcal{A}$,

$$P(a,b|\emptyset) = \frac{e^{u(a)}}{e^{u(a)} + e^{u(b)}}.$$

By (7) and (8), for any $(a',b') \in \mathcal{A}$, the choice probabilities $P(a,b|a',b')$ are represented both by a logit model with utility $u_{(a',b')}$ *and* by a logit model with utility $u^\emptyset_{(a',b')}$. By Lemma

2, these two utility functions differ by a constant. Without loss of generality, we can replace $u_{(a',b')}$ with $u^{\emptyset}_{(a',b')}$ in the family $\{u^h \mid h \in \mathcal{H}^*\}$ without altering (7) and (8). By induction over the length of a history shows that the DSCF can be rationalized by a logit-reinforcement model with reinforcement function R and the prior utility u. *Q.E.D.*

4.4. *Corollaries and Generalizations*

Axiom BCA states that choosing a over b always leads to an increased probability to repeat this choice. Adding this axiom to the previous ones suffices to characterize logit-reinforcement models with *positive* reinforcement.

COROLLARY 1: *A dynamic stochastic choice function $(\bar{\mathcal{H}}, P)$ can be rationalized by a logit model with positive reinforcement learning if and only if it satisfies U-POS, U-LPR, SDA, IDA, HI, and BCA.*

By virtue of the equivalence between Luce and logit models (without updating), it is straightforward to show that U-POS, U-LPR, SDA, IDA, and HI are also necessary and sufficient for a DSCF to be rationalized by reinforcement learning over an arbitrary Luce model with a given, strictly increasing and strictly positive weighting function ω. The same applies for positive reinforcement if axiom BCA is added.

COROLLARY 2: *A DSCF $(\bar{\mathcal{H}}, P)$ can be rationalized by a Luce model with reinforcement learning if and only if it can be rationalized by a logit model with reinforcement learning. Moreover, the reinforcement in the Luce model is positive if and only if it is positive in the logit model.*

Recall that logit models are Fechnerian (Example 2) with the cdf given by $G_{Logit} = 1/(1 + e^{-x})$. The results stated above can be generalized to the class of Fechnerian models for an arbitrary but fixed cdf G, provided it is strictly increasing, as for example the class of probit models. The generalization is immediate if Theorem 1 is restated as follows: a DSCF $(\bar{\mathcal{H}}, P)$ is rationalizable as a logit model with reinforcement if and only if it fulfills SDA, IDA, and HI, and $P(\cdot|h)$ is rationalizable as a logit model for every $h \in \bar{\mathcal{H}}$. It is then straightforward that replacing the cdf does not alter the structure of the arguments.

COROLLARY 3: *Fix a strictly increasing cdf G. A DSCF $(\bar{\mathcal{H}}, P)$ is rationalizable as a Fechner model with cdf G and reinforcement if and only if it fulfills SDA, IDA, HI, and $P(\cdot|h)$ is rationalizable as a Fechner model with cdf G for every $h \in \bar{\mathcal{H}}$. Moreover, reinforcement is positive if and only if, in addition, the DSCF satisfies BCA.*

4.5. *Identification of Logit-Reinforcement Models*

By Lemma 2, if a DSCF can be rationalized by a logit-reinforcement model for some reinforcement function R, the prior utility u is unique up to an additive constant. The following result shows that R is also unique and hence fully identified by the data.

PROPOSITION 2: *Suppose a DSCF can be rationalized by a logit-reinforcement model with reinforcement function R and prior utility u, and also by a logit-reinforcement model with reinforcement function R' and prior utility u'. Then, $R(a) = R'(a)$ for all $a \in A$ and there exists a constant K such that $u'(a) = u(a) + K$ for all $a \in A$.*

PROOF: Suppose a DSCF $(\bar{\mathcal{H}}, P)$ can be rationalized by a logit-reinforcement model with two alternative prior utilities u and u', and two reinforcement functions R and R'. By Lemma 2, there exists a constant K such that $u'(a) = u(a) + K$ for all $a \in A$.

For any $a \in A$, let $b, c \neq a$. Then,

$$\frac{e^{u(a)+R(a)}}{e^{u(a)+R(a)} + e^{u(b)}} = P(a, b|a, c) = \frac{e^{u(a)+K+R'(a)}}{e^{u(a)+K+R'(a)} + e^{u(b)+K}}$$

and hence $u(a) + R(a) - u(b) = u(a) + K + R'(a) - u(b) - K$, implying $R(a) = R'(a)$ for all $a \in A$. *Q.E.D.*

Thus, logit-reinforcement models are *identified* in the sense that both the prior utility u and the reinforcement function R for a given DSCF are unique, up to an additive normalization of the prior utility. Moreover, as verified by inspection of the proof, identification of u relies only on the stochastic choices given no initial history, while identification of R requires only one round of updating. Hence, while the set of all feasible histories $\bar{\mathcal{H}}$ is infinite for the DSCF—and a DSCF can therefore be viewed as generating an infinite dataset of stochastic choice observations—only a finite number of history-dependent stochastic choice observations are needed to identify the model parameters.

5. The RU-Bayesian Model

For the random utility model, Bayesian updating of the underlying stochastic preference is a natural approach to incorporate information from choice observations. We hence now aim to identify the axiomatic properties of the RU-Bayesian model. The problem is challenging for an important conceptual reason, which requires taking a different approach to the axiomatic characterization than for the logit-reinforcement model.

As shown in the previous section, the characterization of the DSCFs that can be rationalized by the logit-reinforcement model follows a step-by-step procedure. First, axioms U-POS and U-LRP characterize the SCFs that can be rationalized by a logit model for any given history of choices, where the utilities are identified uniquely up to addition of a constant (Lemmata 1 and 2). Second, axioms IDA and SDA characterize updating with one new choice observation in terms of a reinforcement of the utilities from the previous step. Moreover, as the proof of Proposition 2 shows, one step of updating identifies a unique reinforcement function. Finally, axiom HI implies that one-step updating fully summarizes the information contained in a history of arbitrary length by ensuring that updated utilities are derived by the accumulation of reinforcements. Given a DSCF that satisfies the axioms, we can thereby identify features of the logit-reinforcement model step by step.

This approach is more challenging for the RU-Bayesian model. There is a well-known counterpart of axioms POS and LRP for the RUM—the Axiom of Revealed Stochastic Preference (ARSP)—which characterizes when a SCF can be rationalized in terms of a distribution over strict preferences. However, unlike the logit model, it is also well-known that RUMs suffer from an identification problem: in general, the same SCF can be rationalized by different distributions over the set of strict preferences. This identification problem is pervasive because, as we show below, simply adding conditional probabilities and requiring Bayesian updating does not always solve the problem, e.g., if only one round of updating is considered. This makes a step-by-step approach to the characterization challenging, since axioms should be expressed in terms of choice probabilities, but choice probabilities do not pin down the underlying parameter of the RUM for the next step of updating.

Intuitively, this identification problem must be resolved with sufficient choice data since, eventually, a single strict preference relation will be identified by the choice data. As a result, our characterization of the RU-Bayesian model is based on a backward-induction logic, which is quite different from the approach to the logit-reinforcement model. In particular, given the aforementioned difficulties with a step-by-step approach in the RU-Bayesian model, it is not sufficient for axiomatic properties to identify the implications of adding choice observations one at a time. Instead, stochastic choices given any history can be related to the stochastic choices for a set of "maximal" histories, each of which identifies a unique strict preference. This relationship can then be used to recover distributions over strict preferences for earlier histories by backward induction, solving the identification problem. Strikingly, our main result in this section shows that a single axiom allows for a full characterization, the Axiom of Bayesian Stochastic Preference (ABSP).

Our presentation of the RU-Bayesian model is, therefore, organized as follows. Subsection 5.1 recals the ARSP, which characterizes SCFs (without updating) that can be rationalized by the RUM. Subsection 5.2 discusses the identification problem for the RUM and provides the intuition for overcoming this issue in our framework. In Subsection 5.3, we then introduce the ABSP and show that it characterizes the DSCFs which can be rationalized by the RU-Bayesian model. Subsection 5.4 discusses implications of the ABSP and illustrates how different properties help discriminate among different classes of models.

5.1. *RUM Stochastic Choice Functions*

Falmagne (1978) and Barberá and Pattanaik (1986) provided the first characterizations of the SCFs that can be rationalized by the RUM. McFadden and Richter (1990) provided an alternative characterization in terms of the now well-known *Axiom of Revealed Stochastic Preference* (ARSP).

AXIOM—ARSP: *For any finite collection of choice problems* $\{(a_i, b_i) \in \mathcal{A} \mid i = 1, \ldots, m\}$, $\sum_{i=1}^{m} P(a_i, b_i) \leq \max_{\succ \in \mathcal{R}} \sum_{i=1}^{m} \mathbb{1}[a_i \succ b_i]$, *where* $\mathbb{1}[\cdot]$ *denotes the indicator function.*

The ARSP asserts that the sum of the choice probabilities corresponding to any finite sequence of binary choice problems cannot exceed the maximal sum that could be induced by some non-stochastic choice function. As such, choice probabilities can always be interpreted as random deviations from non-stochastic choices of some strict preference.

The following Lemma shows that the ARSP characterizes SCFs that can be rationalized by a RUM. The Lemma was proven by McFadden and Richter (1990).

LEMMA 3: *A SCF* P *can be rationalized by the random preference model (hence by the RUM) if and only if it satisfies the ARSP.*

5.2. *Identification in the RU-Bayesian Model*

A fundamental and well-known difficulty of the RUM (e.g., Barberá and Pattanaik, 1986) is that the probability distribution on the set of strict preference profiles representing a SCF (without updating) is not unique, as the following example illustrates.[5] The example we present here makes the additional point that, in some cases, updating can potentially resolve the multiplicity by identifying a unique prior.

[5]This observation also holds if one considers non-binary choices, see e.g. Fischburn (1998, pp. 297–298). A previous example was provided by Barberá and Pattanaik, 1986, Ex. 2.2, which is correct for binary choices but incorrect for non-binary ones, although of course the point they make is correct.

EXAMPLE 6: Let $A = \{a, b, c\}$ and consider two distributions $\pi_1, \pi_2 \in \Delta(\mathcal{R})$ over the set of strict preferences on A, as given in the first two rows of the following table.

Preference	$a \succ b \succ c$	$a \succ c \succ b$	$b \succ a \succ c$	$b \succ c \succ a$	$c \succ a \succ b$	$c \succ b \succ a$
π_1	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
π_2	$1/2$	0	0	0	0	$1/2$
$f_B(\pi_1, (a, b))$	$1/3$	$1/3$	0	0	$1/3$	0
$f_B(\pi_2, (a, b))$	1	0	0	0	0	0

By symmetry, $P_{\pi_1} = P_{\pi_2} = P$ where $P(x, y) = 1/2$ for all $x, y \in A$, $x \neq y$. That is, both distributions generate and represent the same SCF, but they are different elements in the SCM. Suppose now that the choice (a, b) is observed. The last two rows of the table spell out the results of Bayesian updating starting from π_1 or π_2, showing that $f_B(\pi_1, (a, b)) \neq f_B(\pi_2, (a, b))$. That is, while the SCF is not identified, adding a single observation from the DSCF suffices to distinguish the two possible representations.

In the example, adding an observation within a DSCF suffices to identify the correct probability distribution among the two given candidates, and this might suggest that incorporating a round of updating could resolve the identification problem for the RUM. After all, examination of the proof of Proposition 2 in Section 4.5 shows that the presence of a single round of updating already provides full identification in the case of logit-reinforcement models. Unfortunately, this is *not* true for the RU-Bayesian model. That is, a single round of updating is not generally sufficient for identification in the RU-Bayesian case. The following example shows that a given SCF can be represented by two different probability distributions over strict preferences *such that* the respective updated probability distributions given *any* observed choice still induce the same (posterior) choice probabilities. That is, the identification problem persists after updating.

EXAMPLE 7: Let $A = \{a, b, b', c\}$ and consider the two distributions $\pi_1, \pi_2 \in \Delta(\mathcal{R})$ given in the following table (preferences are listed omitting the symbol $\succ$).

Preference	$abb'c$	$ab'bc$	$acbb'$	$acb'b$	$bb'ac$	$b'bac$	$bb'ca$	$b'bca$	$cabb'$	$cab'b$	$cbb'a$	$cb'ba$	all other
π_1	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	0
π_2	$1/6$	0	$1/12$	$1/12$	0	$1/6$	$1/12$	$1/12$	0	$1/6$	$1/6$	0	0

Clearly, $P_{\pi_1} = P_{\pi_2} = P$ where $P(x, y) = 1/2$ for all $(x, y) \in \mathcal{A}$. That is, as in the previous example, both distributions generate and represent the same SCF. However, now suppose the choice (a, b) is observed. Then $\hat{\pi}_1 = f_B(\pi_1, (a, b))$ is uniform over the six preferences in the table above where $a \succ b$. In contrast, $\hat{\pi}_2 = f_B(\pi_2, (a, b))$ places probabilities $1/3, 1/6, 1/6$, and $1/3$ on the preferences $a \succ b \succ b' \succ c$, $a \succ c \succ b \succ b'$, $a \succ c \succ b' \succ b$, and $c \succ a \succ b' \succ b$, respectively. However, $P_{\hat{\pi}_1} = P_{\hat{\pi}_2} = \hat{P}$ with $\hat{P}(a, b) = \hat{P}(a, b') = 1$, $\hat{P}(a, c) = \hat{P}(c, b) = \hat{P}(c, b') = 2/3$, and $\hat{P}(b, b') = 1/2$. That is, after observing (a, b), Bayesian updating yields two different distributions over strict preferences depending on whether the prior was π_1 or π_2, but the induced choice probabilities are identical. Direct calculations show that updated choice probabilities are indeed the same after *any* possible observed choice, and so both priors lead to the same stochastic choice posteriors (see Table C.I in Appendix C). Hence, with a single round of updating, the DSCF still has two different representations in terms of the Bayesian model.

However, adding new choice data does mitigate the problem since, for example, the observation (a, b) at least "reveals" that b is not strictly preferred to a. Intuitively, the identification

problem must, therefore, be resolved with *sufficient* choice data since, eventually, a single, strict preference is identified by elimination. As such, DSCFs incorporate enough information to ensure the following identification result.

PROPOSITION 3: *Suppose a DSCF can be rationalized by the RU-Bayesian model with prior π. Then, it cannot be rationalized by the RU-Bayesian model with a different prior.*

The proof of Proposition 3 will be a by-product of our characterization result in the following section (see Remark 2 below). For now, we only note that identification of the prior distribution over strict preferences can always be achieved with a finite dataset consisting of the stochastic choices for a finite set of maximal histories, which we define next.

5.3. *Characterization of the RU-Bayesian Model*

Our axiomatic characterization of RU-Bayesian model exploits the identification that is eventually achieved in our dynamic framework. We consider a specific subset of *tight maximal* histories that represent the most efficient way to identify strict preferences, and then relate stochastic choices for arbitrary histories to the stochastic choices for tight maximal ones. Somewhat surprisingly, the characterization does not directly use the ARSP but, instead, is based on a single dynamic axiom that establishes the desired relationship of the updated stochastic choices for tight maximal histories.

Let K be the number of alternatives in A, and write this set as $A = \{a_1, \ldots, a_K\}$. A history $h \in \mathcal{H}$ is *tight* if there exists a permutation $\phi : \{1, \ldots, K\} \mapsto \{1, \ldots, K\}$ such that

$$h = \left(\left(a_{\phi(1)}, a_{\phi(2)} \right), \left(a_{\phi(2)}, a_{\phi(3)} \right), \ldots, \left(a_{\phi(n)}, a_{\phi(n+1)} \right) \right),$$

where $n = \ell(h)$ is the length of h. In this case, write $B(h) = a_{\phi(1)}$ ("best alternative") and $W(h) = a_{\phi(n+1)}$ ("worst alternative"). For instance, the history $h = ((a,b),(b,c),(c,d))$ is tight, while the history $h' = ((a,b),(a,c),(a,d))$ is not. The intuition is that the choices in a tight history identify the preference among the involved alternatives, $B(h) \succ a_{\phi(2)} \succ \ldots \succ a_{\phi(n)} \succ W(h)$, and do so in the most efficient way by removing redundancies from the choice data. Thus, each tight history of length $K-1$ identifies exactly one strict preference relation on A, and vice versa. Call a tight history h *maximally tight* if $\ell(h) = K-1$, and denote by $T(\mathcal{H})$ the set of all maximally tight histories.

For a history $h \in \mathcal{H}$ and a pair $(a,b) \in \mathcal{A}$, we write $(a,b) \in h$ if (a,b) is one of the choices observed in h, i.e. $h = (\ldots, (a,b), \ldots)$. We can then define a partial order among histories by writing $h' \subseteq h$ whenever, for all $(a,b) \in \mathcal{A}$, we have that $(a,b) \in h'$ implies $(a,b) \in h$.

For a history $h \in \mathcal{H}$ and distinct $a, b, \in A$, we say that a is revealed preferred to b given h, denoted $a \succ_h b$, if there is a tight history $h^* \subseteq h$ such that $a = B(h^*)$ and $b = W(h^*)$.

Consider a DSCF $(\bar{\mathcal{H}}, P)$. For any history $h \in \bar{\mathcal{H}}$, define the set of maximally tight histories that is *consistent* with h by

$$\mathcal{H}^*(h) = \left\{ h^* \in \bar{\mathcal{H}} \mid h^* \in T(\mathcal{H}) \text{ and, for all } a, b \in A, a \succ_h b \Rightarrow a \succ_{h^*} b \right\}.$$

Note that $\mathcal{H}^*(\emptyset) = \bar{\mathcal{H}} \cap T(\mathcal{H})$, the set of maximally tight histories admissible for the DSCF.

Finally, we want to establish a relationship between stochastic choices for an arbitrary history and the stochastic choices for the set of consistent maximally tight histories. Intuitively, this relationship should be based on the joint probability of a set of choice observations, but joint probabilities are not the primitive for a DSCF. However, we can replicate joint probabilities using conditional probabilities, which do correspond to the data from a DSCF. For any

$h \in \bar{\mathcal{H}}$, $h = (s_1, \ldots, s_n) \neq \emptyset$, define[6] $\Pi(h) = \prod_{i=1}^{n} P(s_i | s_{i+1}, \ldots, s_n)$, where the last term in the product is understood as $P(s_n|\emptyset)$. Define also $\Pi(\emptyset) = 1$. The quantity $\Pi(h)$ captures the probability of sequentially observing the choices listed in h according to the DSCF, and hence is defined on primitives. The definition of a DSCF implies that $\Pi(h) > 0$ for all $h \in \bar{\mathcal{H}}$. Our characterization relies on the relation between such probabilities for arbitrary histories and the corresponding probabilities for consistent maximally tight histories, as captured in the following *Axiom of Bayesian Stochastic Preference*.

AXIOM—ABSP: *For all* $h \in \bar{\mathcal{H}}$, $\Pi(h) = \sum_{h^* \in \mathcal{H}^*(h)} \Pi(h^*)$.

The ABSP has an obviously Bayesian flavor. It requires that the chain (product) of updated probabilities along a given history h add up to the sum of the analogous products along all maximal but "efficient" (i.e., tight) histories that the observations in h can give rise to by sequentially adding choices which do not contradict those in h. For example, suppose $A = \{a, b, c\}$ and consider the history $h = (a, b)$, so that $\Pi(h) = P(a, b|\emptyset)$. The ABSP implies $P(a, b|\emptyset) = P(a, b|b, c)P(b, c|\emptyset) + P(c, a|a, b)P(a, b|\emptyset) + P(a, c|c, b)P(c, b|\emptyset)$, taking into account the three possible maximally tight histories (each of them *ex post* identifying an underlying preference).

The name ABSP will be justified after the fact since the characterization below, together with Lemma 3, implies that the SCF $P(\cdot|h)$ must fulfill the ARSP for every history $h \in \bar{\mathcal{H}}$ (see Section 5.4 below).

THEOREM 2: *A dynamic stochastic choice function* $(\bar{\mathcal{H}}, P)$ *can be rationalized by the RU-Bayesian model if and only if it satisfies the ABSP.*

PROOF: Necessity is straightforward, as the ABSP then reduces to application of Bayes' rule. To show sufficiency, first note that applying the ABSP to the empty history yields $1 = \Pi(\emptyset) = \sum_{h^* \in \mathcal{H}^*(\emptyset)} \Pi(h^*)$, and $\mathcal{H}^*(\emptyset)$ is the set of all maximally tight histories which are feasible in the DSCF. Obviously, for every maximally tight history $h \in T(\mathcal{H})$, there is one and only one strict preference on A, $\succ$, such that $\succ_h = \succ$. Hence, we can define an SCF within the RUM by setting $\pi(\succ) = \Pi(\succ_h)$ if there exists an $h \in \mathcal{H}^*(\emptyset)$ such that $\succ = \succ_h$, and $\pi(\succ) = 0$ if not. Consider now any $s = (a, b) \in \mathcal{A}$. Suppose first $s \in \bar{\mathcal{H}}$. A maximally tight $h \in \mathcal{H}^*(\emptyset)$ is in $\mathcal{H}^*(s)$ if and only if $a \succ_h b$. By the ABSP,

$$P(s|\emptyset) = \Pi(s) = \sum_{h^* \in \mathcal{H}^*(s)} \Pi(h^*) = \sum \{\pi(\succ) \mid \succ \in \mathcal{R},\ a \succ b\}.$$

If $s \notin \bar{\mathcal{H}}$, this means $P(s|\emptyset) = 0$ and $\mathcal{H}^*(s) = \emptyset$, hence the last equality holds trivially. Thus, the RUM rationalizes the probabilities $P(s|\emptyset)$.

Consider now any history $h \in \bar{\mathcal{H}}$, and let $s = (a, b) \in \mathcal{A}$ such that $s \circ h \in \bar{\mathcal{H}}$. By definition, $\Pi(s \circ h) = P(s|h) \cdot \Pi(h)$, and applying the ABSP yields

$$P(s|h) = \frac{\Pi(s \circ h)}{\Pi(h)} = \frac{\sum \{\Pi(h^*) \mid h^* \in \mathcal{H}^*(s \circ h)\}}{\sum \{\Pi(h^*) \mid h^* \in \mathcal{H}^*(h)\}}$$

[6] We abuse notation slightly by writing $P(s_1|s_2, \ldots, s_n)$ instead of $P(s_1|(s_2, \ldots, s_n))$.

$$= \frac{\sum \{\pi(\succ) \mid a \succ b \text{ and } c \succ d \text{ for all } (c,d) \in h\}}{\sum \{\pi(\succ) \mid c \succ d \text{ for all } (c,d) \in h\}}$$

as required by Bayesian updating given the original SCF. If $h \in \bar{\mathcal{H}}$ but $s \circ h \notin \bar{\mathcal{H}}$, then $P(s|h) = 0$, and the numerator in this expression (but not the denominator) is also zero. Hence, the DSCF is rationalized by the RU-Bayesian model. *Q.E.D.*

REMARK 2: The proof of Theorem 2 clarifies the identification result in Proposition 3. If the DSCF $(\bar{\mathcal{H}}, P)$ can be rationalized by the RU-Bayesian model, then the prior distribution over strict preferences π is defined by the condition that, for each $\succ \in \mathcal{R}$, $\pi(\succ) = \Pi(h)$ for the unique tight maximal history such that $\succ_h = \succ$. The set of tight maximal histories is finite, and so a finite collection of history-dependent SCFs from the DSCF suffices to identify a unique prior. Thus, the identification problem is fully resolved in our framework.

5.4. *The ARSP and the ABSP*

For the RU-Bayesian model, the SCF following any history can be rationalized by the RUM, and this property of the SCF is characterized by the ARSP. As such, the following axiom is an immediate implication of the ABSP (similar to Axioms U-POS and U-LRP in the context of the logit-reinforcement model).

AXIOM—U-ARSP: *For all $h \in \bar{\mathcal{H}}$, the SCF $P(\cdot|h)$ satisfies ARSP.*

By Lemma 3, it is clear that a DSCF can be rationalized by the RU-Bayesian model only if U-ARSP is satisfied.[7] However, the distinctive implications of the RU-Bayesian model arise because Bayesian updating imposes additional discipline across the updated SCFs following different choice observations. The ABSP formalizes the relationship across SCF given different histories by establishing a relationship with the complete set of stochastic choices given maximally tight histories.

For instance, Axiom U-ARSP does not distinguish between DSCFs that can be rationalized by the RU-Bayesian model and DSCFs that can be rationalized by a logit-reinforcement model, since every SCF with choice probabilities defined by a logit function can also be rationalized by the RUM (Block and Marschak, 1960), hence must also satisfy the ARSP. However, while the history-dependent SCF generated by logit-reinforcement model satisfy the ARSP, they are not characterized by the ARSP (which is necessary but not sufficient for Axioms POS and LPR). The following example illustrates a SCM-U that, in terms of the properties of the SCF following any given history, is indistinguishable from the RU-Bayesian model, and yet is not observationally equivalent to the RU-Bayesian model in terms of DSCFs because it does imply the ABSP.

EXAMPLE 8—Random-RU-Bayesian Model: As in the RU-Bayesian model, let $\Delta(\mathcal{R})$ be the set of probability distributions on the set of strict preferences $\mathcal{R}$ for the set of alternatives A. Let $\Theta = \Delta^F(\Delta(\mathcal{R}))$ be the set of simple probability distributions on $\Delta(\mathcal{R})$, which can be interpreted as the possible beliefs of an observer about the random preferences of a decision maker. That is, $\theta \in \Theta$ is a probability distribution with finite support over $\Delta(\mathcal{R})$.[8]

[7] Further implications of the ABSP are discussed in Appendix D.

[8] We focus on probability distributions with finite support to avoid introducing new notation at this stage.

For $\theta \in \Theta$, let $\theta(\pi)$ denote the probability of the distribution over strict preferences $\pi \in \Delta(\mathcal{R})$, and let supp θ be the support of $\theta \in \Theta$. We then define $P_\theta(a,b) = \sum_{\pi \in \text{supp } \theta} P_\pi(a,b)\theta(\pi)$ where $P_\pi(a,b) = \sum_{\succ \in \mathcal{R}: a \succ b} \pi(\succ)$ as in the RU-Bayesian model.

Given some prior parameter $\theta_0 \in \Theta$, reflecting the prior beliefs of the observer, there is a strictly positive probability of observing the choice (a,b) if there is a random preference π in the support of θ that assigns a strictly positive probability to this choice observation. The set of admissible one-period histories is therefore $\bar{\mathcal{A}} = \{(a,b) : \mathcal{A} : \exists \pi \in \text{supp } \theta_0 \text{ with } P_\pi(a,b) > 0\}$. Confronted with a new choice observation $s \in \bar{\mathcal{A}}$, the observer updates beliefs using Bayes' rule, and so

$$f_{RB}(\theta, s) = \begin{cases} \theta & \text{if } s = \emptyset \\ \theta_s & \text{if } s \neq \emptyset \end{cases},$$

where θ_s is defined by $\theta_s(\pi) = \frac{P_\pi(s)\theta_0(\pi)}{\sum_{\pi' \in \text{supp } \theta_0} P_{\pi'}(s)\theta_0(\pi')}$.

The tuple $(\Theta, \{P_\theta\}_{\theta \in \Theta}, f_{RB})$ is a SCM-U, which we call the *random-RU-Bayesian model*. Given the prior θ_0, the corresponding DSCF following history h is described by $P(a,b|h) = \sum_{\pi \in \text{supp } \theta_h} P_\pi(a,b)\theta_h(\pi)$, where $\theta_h = f_{RB}(\theta_0, h)$.

Note that $P_\pi(a,b)$ is not updated for the history h. If we were to replace $P_\pi(a,b)$ with $P_\pi(a,b|h)$ we would obtain a model that is observationally equivalent to the RU-Bayesian model, in which the prior $\pi_0 \in \Delta(\mathcal{R})$ is defined by $\pi_0(\succ) \equiv \sum_{\pi \in \text{supp } \theta_0} \pi(\succ)\theta_0(\pi)$. By contrast, in the random-RU-Bayesian model, the decision maker has a *random preference*, rather than a deterministic strict preference $\succ$. The random preference is unknown to the observer and, confronted with a new choice observation, the observer updates prior beliefs θ_0 over the set of random preferences.

The random-RU-Bayesian model retains the flavor of Bayesian updating. Similarly to the RU-Bayesian, any DSCF rationalized by the random-RU-Bayesian model satisfies U-ARSP. Indeed, it is straightforward to show that, following any *given* history, a SCF can be rationalized by the RUM if and only if it can be rationalized by the random-RUM model. For a fixed history, the random-RU-Bayesian model and RU-Bayesian model, therefore, generate stochastic choices that are indistinguishable.

However, in general, the random-RU-Bayesian model makes different predictions across the updated SCFs following new choice observations. A DSCF rationalized by the RU-Bayesian model with prior $\pi_0 \in \Delta(\mathcal{R})$ can also be rationalized by the random-RU-Bayesian model with a prior $\theta_0 \in \Delta(\Delta(\mathcal{R}))$ given by $\theta_0(\delta_\succ) = \pi_0(\succ)$ for all $\succ \in \mathcal{R}$, where $\delta_\succ \in \Delta(\mathcal{R})$ is the Dirac distribution with probability 1 on the strict preference $\succ$. As such, the DSCFs that can be rationalized by the RU-Bayesian model are a subset of the DSCFs that can be rationalized by the random-RU-Bayesian model. However, the inclusion is strict because a DSCF that is rationalized by the random-RU-Bayesian model can *only* be rationalized by the RU-Bayesian model if the support of θ_0 contains only Dirac distributions.

To see why, first observe that a DSCF can be rationalized by the RU-Bayesian model only if, for any choice observation $s \in \mathcal{A}$ and history h such that $s \circ h$ is admissible, $P(s|s \circ h) = 1$. That is, if alternative a is ever observed being chosen over b, any predicted future choice from $\{a,b\}$ becomes deterministic. Therefore, we call this implication of the RU-Bayesian model *determinism*. The determinism property of the RU-Bayesian model is immediate from the representation but can be derived directly from the ABSP: consider any admissible history h and choice s such that $P(s|h) > 0$. Since h is admissible, it follows by definition that $\Pi(h) > 0$. Moreover, the histories $s \circ h$ and $s \circ s \circ h$ are consistent with exactly the same

set of maximally tight histories, i.e., $\mathcal{H}^*(s \circ h) = \mathcal{H}^*(s \circ s \circ h)$. Therefore, the ABSP implies $P(s|s \circ h)P(s|h)\Pi(h) = P(s|h)\Pi(h)$, and hence $P(s|s \circ h) = 1$.

Now consider a DSCF rationalized by the random-RU-Bayesian model with prior θ_0 where $\pi(\succ), \pi(\succ') > 0$ for some $\pi \in \text{supp}\theta_0$. If $\succ \neq \succ'$, there is some pair of alternatives $a, b \in A$ such that $a \succ b$ and $b \succ' a$. As a result, $P_\pi(a, b) \in (0, 1)$ and therefore

$$P_{f(\theta_0,(a,b))}(a,b) = \frac{\sum_{\pi' \in \text{supp}\theta_0} P_{\pi'}(a,b)^2 \theta_0(\pi')}{\sum_{\pi'' \in \text{supp}\theta_0} P_{\pi''}(a,b)\theta_0(\pi'')} \neq 1.$$

Therefore, a DSCF rationalized by the random-RU-Bayesian model with prior θ_0 satisfies determinism if and only if the support of θ_0 is concentrated on Dirac distributions over strict preferences.[9]

6. Preference Change

Our framework can also be used to model preference change in the presence of stochastic choice. This provides a link to a large literature in psychology, neuroscience, and cognitive science, which suggests that "attitudes," a construct close to the economic idea of preferences, are not stable. Widespread evidence going back to Brehm (1956) suggests that the mere act of choice can feed back into and alter pre-existing attitudes: people adjust to "spread out" their self-reported evaluations, typically adjusting the evaluation of chosen options up (see, e.g., Harmon-Jones and Mills, 1999, Ariely and Norton, 2008).[10] A common explanation involves cognitive dissonance (Festinger, 1957, Joule, 1986), whereby attitudes and beliefs are brought in line with actions after the fact.

Choice-induced preference change lends itself to a formalization in terms of SCM-U. However, the translation of the statement into specific properties shows that some aspects may be neither surprising nor particularly removed from "rationalistic" models. Consider, for instance, axiom BCA, which asserts that the probability of choosing a over b should increase if this choice has been previously made. As shown in Theorem 1 and Corollary 1, this property is *not* always satisfied by the logit-reinforcement model, but is a characterizing property of *positive* reinforcement. That is, BCA merely captures positive updating of decision values for non-aversive options. Further, as discussed in Section 5.4, the ABSP implies that $P(a, b|(a, b) \circ h) = 1$, hence axiom BCA is always trivially satisfied for a DSCF that can be rationalized by an RU-Bayesian model.

This BCA property, however, only captures one aspect of choice-induced attitude change as discussed in the literature from psychology and neuroscience. A closer formalization of "preference change" is as follows.

AXIOM—PC: *For all* $a, b, c \in \mathcal{A}$ *with* $a \neq b, c$ *and all* $h \in \bar{\mathcal{H}}$ *such that* $0 < P(a, c|h) < 1$, $P(a, c|(a, b) \circ h) > P(a, c|h)$.

[9] A different model fulfilling Bayesian principles but violating determinism would be a trembling-hand model where the decision maker has a deterministic preference but makes mistakes with a probability ε, and the modeler uses Bayes' rule while taking mistakes into account. This model is defined formally in Appendix D.

[10] The evidence is controversial (Fudenberg et al., 2012, Maniadis et al., 2014). Psychological paradigms have been shown to exhibit a statistical bias that can result in apparent preference change even if participants have immutable preferences (Chen and Risen, 2010, Izuma and Murayama, 2013, Alós-Ferrer and Shi, 2015), although results have been reestablished with improved designs (Sharot et al., 2010, Alós-Ferrer et al., 2012).

Axiom PC states that the probability of choosing an option increases after it has been chosen, *independently of which alternative it is paired with.* Obviously, the RU-Bayesian model does not, in general, exhibit this property. Suppose, for instance, that $A = \{a, b, c\}$ and π puts probability $1/2$ each on the preferences $c \succ a \succ b$ and $b \succ a \succ c$. Then, $P(a, c|\emptyset) = 1/2$ but $P(a, c|a, b) = 0$, violating PC. The DSCF generated by a logit model with positive reinforcement fulfills PC, because a positive reinforcer $R(a) > 0$ results in larger updated utility $u(a) + R(a)$ and hence increased choice probability independently of which other option is available, but the implication fails without positive reinforcement. Hence, the logit model with positive reinforcement captures and reflects evidence on choice-induced attitude change, but the other models considered here preclude this phenomenon. Indeed, in the presence of U-POS, U-LPR, HI and SDA, the two axioms PC and BCA are equivalent: on one hand, PC encompasses BCA as a particular case ($c = b$ in the axiom's statement); on the other hand, if BCA is fulfilled in the presence of the other axioms, Corollary 1 shows that the DSCF can be rationalized by a logit model with positive reinforcement, which in turn implies that the DSCF satisfies PC.

PROPOSITION 4: *A DSCF satisfies U-POS, U-LPR, SDA, HI and Axiom PC if and only if it can be rationalized by a logit model with positive reinforcement learning.*

In this sense, our approach suggests that logit models with positive reinforcement learning are an appropriate formal framework to study (and test) the choice-induced preference change phenomenon. Interestingly, our approach also sheds light on some long-standing questions in the literature. For instance, *self-perception theory* advocates an alternative interpretation of choice-induced attitude change, which differs fundamentally from interpretations based on cognitive dissonance: the decision maker is assumed to be (consciously or not) unsure about his or her own preferences, and uses their own observed choices as signals (Bem, 1967a,b). The claim is that, in the face of these observations, chosen options are reevaluated upwards, but this implication is not typically formalized. Our analysis shows that, to imply PC, this theory must be complemented with a behavioral component e.g. updating of a decision value through positive reinforcement. If, for instance, self-perception theory is formalized in a rationalistic fashion (where the decision maker learns as an external, rational observer would), the decision maker would hold beliefs on their own preferences and update them following Bayes' rule, in which case the model becomes equivalent to the RU-Bayesian model, which, as we have seen, does not satisfy axiom PC.

7. Discussion

The problem of updating a model that predicts an agent's decisions in response to new choice information is pervasive in economics, psychology, and neuroscience. Normative economics typically adopts a Bayesian approach, where beliefs on an appropriate space are updated through Bayes' rule. Decision neuroscience has identified decision values in the brain and obtained widespread evidence that such values are updated following reinforcement processes. Applied economics and microeconometrics typically consider random utility models, which are equivalent to the first class in terms of static choices, but provide a utility function which could be updated as in the second class. We show that those and many other models are instances of a general, unified framework which jointly addresses how choices are derived from an underlying (functional) parameter space, and how those parameters are updated. Once this is done, the connections between the various approaches become apparent, and the structural assumptions underlying each model can be identified.

Those structural assumptions, however, condition how new information affects the posterior. It is well-known that the same choice probabilities can be explained either through a probability distribution over deterministic preferences (a random preference model) or through a fixed utility function affected by noise (a random utility model). A Bayesian modeler would, however, focus on updating the beliefs, while neuroscience suggests that the human brain relies on certain, cardinal decision values and updates them, very much as if the utility function in a random utility model were updated. Starting with the same initial probabilities and adding the same new choice evidence, both approaches will typically end up with different predictions. Through our characterizations, we identify the structural assumptions behind the most prominent stochastic choice models, which opens the door to model separation and testing. For instance, we provide an approach to distinguish the value-based reinforcement approach and the belief-based preference approach.

Each characterization we develop is also of independent interest. The "neuroeconomic" logit-reinforcement model is characterized by simple properties, which essentially make clear that only the values of the chosen options are updated, with no indirect inferences made. Positive reinforcement is not a structural part of the model but can easily be added, resulting in a subclass which, interestingly, is characterized by a property reflecting choice-induced attitude change, a phenomenon implied by cognitive dissonance and widely discussed in psychology. The RUM suffers from identification problems which subsist after a round of updating, in the sense that the same choice probabilities *even after every possible updating* can be explained by different prior beliefs. However, those problems are solved in our framework, where histories of arbitrary length are allowed, and the resulting characterization relies on a single property, the Axiom of Bayesian Stochastic Preference.

Although we have focused on the two most-relevant models, our framework goes well beyond those. A myriad of model variants spring to mind, ranging from Bayesian updating while allowing for errors or inherently noisy behavioral types to complex forms of reinforcement where the discarded option is discounted or influences the value of the reinforcer. For each, we provide a framework to identify the structural differences with previous models, and hence the appropriate behavioral tests to separate and identify them. At a conceptual level, our contribution provides a common language to the study the observational implications of such models, and shows that dialects of this language are already spoken both by Bayesian modelers and neuroeconomists.

REFERENCES

ALÓS-FERRER, C., E. FEHR, AND N. NETZER (2020): "Time Will Tell: Recovering Preferences when Choices are Noisy," *Journal of Political Economy*, forthcoming.

ALÓS-FERRER, C., D.-G. GRANIĆ, F. SHI, AND A. K. WAGNER (2012): "Choices and Preferences: Evidence from Implicit Choices and Response Times," *Journal of Experimental Social Psychology*, 48, 1336–1342.

ALÓS-FERRER, C. AND F. SHI (2015): "Choice-Induced Preference Change and the Free-Choice Paradigm: A Clarification," *Judgment and Decision Making*, 10, 34–49.

ANDERSON, S. P., J.-F. THISSE, AND A. DE PALMA (1992): *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: MIT Press.

ARIELY, D., G. LOEWENSTEIN, AND D. PRELEC (2003): ""Coherent Arbitrariness": Stable Demand Curves without Stable Preferences," *Quarterly Journal of Economics*, 118, 73–105.

ARIELY, D. AND M. I. NORTON (2008): "How Actions Create – Not Just Reveal – Preferences," *Trends in Cognitive Sciences*, 12, 13–16.

BALDASSI, C., S. CERREIA-VIOGLIO, F. MACCHERONI, AND M. MARINACCI (2020): "A Behavioral Characterization of the Drift Diffusion Model and its Multi-Alternative Extension to Choice under Time Pressure," *Management Science*, 66, 5075–5093.

BALLESTA, S., W. SHI, K. E. CONEN, AND C. PADOA-SCHIOPPA (2020): "Values Encoded in Orbitofrontal Cortex Are Causally Related to Economic Choices," *Nature*, 588, 450–453.

BARBERÁ, S. AND P. K. PATTANAIK (1986): "Falmagne and the Rationalizability of Stochastic Choices in Terms of Random Orderings," *Econometrica*, 54, 707–715.

BELLEMARE, C., S. KRÖGER, AND A. VAN SOEST (2008): "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica*, 76, 815–839.

BEM, D. J. (1967a): "Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena," *Psychological Review*, 74, 183–200.

——— (1967b): "Self-Perception: The Dependent Variable of Human Performance," *Organizational Behavior and Human Performance*, 2, 105–121.

BLOCK, H. D. AND J. MARSCHAK (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. by I. Olkin, Stanford: Stanford University Press, 97–132.

BREHM, J. W. (1956): "Postdecision Changes in the Desirability of Alternatives," *The Journal of Abnormal and Social Psychology*, 52, 384–389.

BROCAS, I. AND J. D. CARRILLO (2012): "From Perception To Action: An Economic Model of Brain Processes," *Games and Economic Behavior*, 75, 81–103.

BRUHIN, A., E. FEHR, AND D. SCHUNK (2018): "The Many Faces of Human Sociality: Uncovering the Distribution and Stability of Social Preferences," *Journal of the European Economic Association*, 17, 1025–1069.

BRUHIN, A., H. FEHR-DUDA, AND T. EPPER (2010): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica*, 78, 1375–1412.

BUSH, R. R. AND F. MOSTELLER (1951): "A Mathematical Model for Simple Learning," *Psychological Review*, 58, 312–323.

——— (1955): *Stochastic Models for Learning*, New York: Wiley.

CAPLIN, A. AND M. DEAN (2008): "Dopamine, Reward Prediction Error, and Economics," *Quarterly Journal of Economics*, 123, 663–701.

——— (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105, 2183–2203.

CAPLIN, A., M. DEAN, P. W. GLIMCHER, AND R. B. RUTLEDGE (2010): "Measuring Beliefs and Rewards: A Neuroeconomic Approach," *Quarterly Journal of Economics*, 125, 923–960.

CAPLIN, A. AND D. MARTIN (2015): "A Testable Theory of Imperfect Perception," *Economic Journal*, 125, 18–202.

CHEN, M. K. AND J. L. RISEN (2010): "How Choice Affects and Reflects Preferences: Revisiting the Free-Choice Paradigm," *Journal of Personality and Social Psychology*, 99, 573–594.

CONTE, A., J. D. HEY, AND P. G. MOFFATT (2011): "Mixture Models of Choice Under Risk," *Journal of Econometrics*, 162, 79–88.

COSTA-GOMES, M., V. P. CRAWFORD, AND B. BROSETA (2001): "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 69, 1193–1235.

CROSS, J. G. (1973): "A Stochastic Learning Model of Economic Behavior," *Quarterly Journal of Economics*, 87, 239–266.

——— (1983): *A Theory of Adaptive Economic Behavior*, Cambridge: Cambridge University Press.

DAW, N. D. AND P. N. TOBLER (2014): "Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning," in *Neuroeconomics: Decision Making and the Brain*, ed. by P. W. Glimcher and E. Fehr, London: Academic Press, 283–298, 2nd ed.

DEBREU, G. (1958): "Stochastic Choice and Cardinal Utility," *Econometrica*, 26, 440–444.

EL-GAMAL, M. A. AND D. M. GRETHER (1995): "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*, 90, 1137–1145.

FALMAGNE, J.-C. (1978): "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology*, 18, 52–72.

FESTINGER, L. (1957): *A Theory of Cognitive Dissonance*, Stanford, CA: Stanford University Press.

FISCHBURN, P. C. (1998): "Stochastic Utility," in *Handbook of Utility Theory*, ed. by S. Barberà, P. J. Hammond, and C. Seidl, Kluwer Academic Publishers, vol. 1: Principles, chap. 7, 273–319.

FRICK, M., R. IIJIMA, AND T. STRZALECKI (2019): "Dynamic Random Utility," *Econometrica*, 87, 1941–2002.

FUDENBERG, D., D. K. LEVINE, AND Z. MANIADIS (2012): "On the Robustness of Anchoring Effects in WTP and WTA Experiments," *American Economic Journal: Microeconomics*, 4, 131–145.

FUDENBERG, D., P. STRACK, AND T. STRZALECKI (2018): "Speed, Accuracy, and the Optimal Timing of Choices," *American Economic Review*, 108, 3651–3684.

FUDENBERG, D. AND T. STRZALECKI (2015): "Dynamic Logit with Choice Aversion," *Econometrica*, 83, 651–691.

GUL, F. AND W. PESENDORFER (2006): "Random Expected Utility," *Econometrica*, 74, 121–146.

HARMON-JONES, E. AND J. E. MILLS (1999): *Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology*, Washington, DC: American Psychological Association.

HOLROYD, C. B. AND M. G. COLES (2002): "The Neural Basis of Human Error Processing: Reinforcement Learning, Dopamine, and the Error-Related Negativity," *Psychological Review*, 109, 679–709.

IZUMA, K. AND K. MURAYAMA (2013): "Choice-Induced Preference Change in the Free-Choice Paradigm: A Critical Methodological Review," *Frontiers in Psychology*, 4, 1–12.

JOULE, R. V. (1986): "Twenty Five On: Yet Another Version of Cognitive Dissonance Theory?" *European Journal of Social Psychology*, 16, 65–78.

LUCE, R. D. (1959): *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley.

——— (1977): "The Choice Axiom after Twenty Years," *Journal of Mathematical Psychology*, 15, 215–233.

LUCE, R. D. AND P. SUPPES (1965): "Preference, Utility and Subjective Probability," in *Handbook of Mathematical Psychology, Vol. 3*, ed. by R. D. Luce, R. R. Bush, and E. Galanter, New York: John Wiley & Sons, 249–410.

MANIADIS, Z., F. TUFANO, AND J. A. LIST (2014): "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects," *American Economic Review*, 104, 277–290.

MARSCHAK, J. (1960): "Binary Choice Constraints on Random Utility Indicators," in *Stanford Symposium on Mathematical Methods in the Social Sciences*, ed. by K. J. Arrow, Stanford, CA: Stanford University Press, 312–329.

MCFADDEN, D. L. (2001): "Economic Choices," *American Economic Review*, 91, 351–378.

MCFADDEN, D. L. AND M. K. RICHTER (1990): "Stochastic Rationality and Revealed Preference," in *Preferences, Uncertainty, and Optimality: Essays in Honor of Leonid Hurwicz*, ed. by J. S. Chipman, D. L. McFadden, and M. K. Richter, Boulder, Colorado: Westview Press, 163–186.

MOFFATT, P. G. (2015): *Experimetrics: Econometrics for Experimental Economics*, London: Palgrave Macmillan.

PADOA-SCHIOPPA, C. AND J. A. ASSAD (2006): "Neurons in the Orbitofrontal Cortex Encode Economic Value," *Nature*, 441, 223–226.

PLATT, M. L. AND P. W. GLIMCHER (1999): "Neural Correlates of Decision Variables in Parietal Cortex," *Nature*, 400, 233–238.

RATCLIFF, R. (1978): "A Theory of Memory Retrieval," *Psychological Review*, 85, 59–108.

SCHULTZ, W. (1998): "Predictive Reward Signal of Dopamine Neurons," *Journal of Neurophysiology*, 80, 1–27.

——— (2010): "Dopamine Signals for Reward Value and Risk: Basic and Recent Data," *Behavioral and Brain Functions*, 6, 1–9.

——— (2013): "Updating Dopamine Reward Signals," *Current Opinion in Neurobiology*, 23, 229–238.

SCHULTZ, W., P. DAYAN, AND P. R. MONTAGUE (1997): "A Neural Substrate of Prediction and Reward," *Science*, 275, 1593–1599.

SHADLEN, M. N. AND R. KIANI (2013): "Decision Making as a Window on Cognition," *Neuron*, 80, 791–806.

SHADLEN, M. N. AND D. SHOHAMY (2016): "Decision Making and Sequential Sampling from Memory," *Neuron*, 90, 927–939.

SHAROT, T., C. M. VELASQUEZ, AND R. J. DOLAN (2010): "Do Decisions Shape Preference? Evidence from Blind Choice," *Psychological Science*, 21, 1231–1235.

SUTTON, R. S. AND A. G. BARTO (1998): *Reinforcement Learning: An Introduction*, Cambridge, MA: MIT Press.

APPENDICES

APPENDIX A: Omitted Proofs

A.1. *Proof of Lemma 1*

First, we show that POS and LPR are necessary. Fix any strictly increasing $\omega : \mathbb{R} \mapsto \mathbb{R}_{++}$ and $u \in \mathcal{U}$. Since ω has strictly positive values, $P_u^\omega(a,b) > 0$ for any $(a,b) \in \mathcal{A}$, and so P_u^ω satisfies POS. To show that LPR holds, let $a, b, c \in \mathcal{A}$ be three distinct alternatives. Then,

$$\frac{P_u^\omega(a,b)}{P_u^\omega(b,a)}\frac{P_u^\omega(b,c)}{P_u^\omega(c,b)} = \frac{\omega(u(a))}{\omega(u(b))}\frac{\omega(u(b))}{\omega(u(c))} = \frac{\omega(u(a))}{\omega(u(c))} = \frac{P_u^\omega(a,c)}{P_u^\omega(c,a)},$$

and it follows immediately that $P_u^\omega(a,b)P_u^\omega(b,c)P_u^\omega(c,a) = P_u^\omega(a,c)P_u^\omega(c,b)P_u^\omega(b,a)$, establishing LPR.

Now we show that POS and LPR are sufficient. Let P be a SCF that satisfies POS and LPR. Fix an arbitrary alternative $c \in \mathcal{A}$, and define $u(c) = 0$. For $a \in \mathcal{A} \setminus \{c\}$, let $u(a) = \ln\left(\frac{P(a,c)}{P(c,a)}\right)$, which is well-defined because P satisfies POS. Now consider any distinct alternatives $a, b \in A \setminus \{c\}$. By definition, $e^{u(a)} = \frac{P(a,c)}{P(c,a)}$, from which it follows that

$$P(a,c) = \frac{e^{u(a)}}{e^{u(a)}+1} = \frac{e^{u(a)}}{e^{u(a)}+e^{u(c)}}.$$

Since P satisfies POS and LPR, a direct computation shows

$$\frac{P(a,b)}{P(b,a)} = \frac{\dfrac{P(a,c)}{P(c,a)}}{\dfrac{P(b,c)}{P(c,b)}}.$$

and using $P(b,a) = 1 - P(a,b)$ we obtain

$$P(a,b) = \frac{\dfrac{P(a,c)}{P(c,a)}}{\dfrac{P(a,c)}{P(c,a)} + \dfrac{P(b,c)}{P(c,b)}} = \frac{e^{u(a)}}{e^{u(a)}+e^{u(b)}}.$$

Hence, P can be rationalized as a logit model (and hence as a Luce model).

A.2. *Proof of Lemma 2*

Suppose P can be rationalized by the logit model with either u or u' as parameter. Fix $b \in A$ and define $K = u'(b) - u(b)$. Then, for any $a \in A$,

$$\frac{e^{u(a)}}{e^{u(a)}+e^{u(b)}} = P(a,b|\emptyset) = \frac{e^{u'(a)}}{e^{u'(a)}+e^{u'(b)}}$$

which implies $u(a) - u(b) = u'(a) - u'(b)$ or $u'(a) = u(a) + K$ for all a.

A.3. *Proof of Corollary 1*

It is obvious that BCA is satisfied in the case of positive reinforcement. To prove sufficiency, proceed as in Step 1 of the proof of Theorem 1 to identify $R^h(a) = S(a,b,c)$ for any $a \in A$ and $h \in \mathcal{H}$. We claim that, under BCA, $S(a,b,c) > 0$ in Step 1a. For, by BCA, for all $(a,b) \in \mathcal{A}$ with $P^h(a,b) > 0$, $P^h(a,b|a,b) > P^h(a,b|\emptyset)$. Hence, by Equation (6),

$$u^h_{a,b}(b) - u^h_{a,b}(a) < u^h(b) - u^h(a)$$

implying that

$$S(a,b,b) = \left[u^h_{a,b}(a) - u^h_{a,b}(b)\right] - \left[u^h(a) - u^h(b)\right] > 0.$$

Since, by Step 1a, $S(a,b,c) = S(a,b,b)$ for all $c \neq a$, we obtain that $R^h(a) = S(a,b,c) > 0$, as claimed. The remaining argument then follows the proof of Theorem 1.

A.4. *Proof of Corollary 2*

Let $(\bar{\mathcal{H}}, P)$ be a DSCF that can be rationalized by a Luce model with reinforcement learning: for some strictly increasing $\omega : \mathbb{R} \mapsto \mathbb{R}_{++}$, $u^* \in \mathcal{U}$ and function $R^* : A \to \mathbb{R}$,

$$P^h(a,b|s) = P^{\omega}_{u^*_s}(a,b) = \frac{\omega(u^*_s(a))}{\omega(u^*_s(a)) + \omega(u^*_s(b))} \quad \forall s \in \mathcal{A}^*,\ h \in \bar{\mathcal{H}}, \text{ and } (a,b) \in \mathcal{A},$$

(recall Section 3.3) where $u^*_s(a) = u^*(a)$ when $s = \emptyset$ or $s = (c,d)$ and $c \neq a$, and $u^*_s(a) = u^*(a) + R^*(a)$ when $s = (c,d)$ and $c = a$.

For any $a \in A$, define $u(a) = \ln(\omega(u^*(a)))$ and $R(a) = \ln(\omega(u^*(a) + R^*(a))) - \ln(\omega(u^*(a)))$. Then, for $a \neq b, c$,

$$P(a,b|\emptyset) = P(a,b|c,d) = \frac{\omega(u^*(a))}{\omega(u^*(a)) + \omega(u^*(b))} = \frac{e^{u(a)}}{e^{u(a)} + e^{u(b)}}$$

$$P(a,b|a,c) = \frac{\omega(u^*(a) + R^*(a))}{\omega(u^*(a) + R^*(a)) + \omega(u^*(b))} = \frac{e^{u(a)+R(a)}}{e^{u(a)+R(a)} + e^{u(b)}}.$$

Hence, the DSCF $P(\cdot|\cdot)$ can also be rationalized by a logit model with utility function u^* and the reinforcement function $R^* : A \to \mathbb{R}$.

The converse implication is obtained by constructing $u^*(a) = \omega^{-1}\left(e^{u(a)}\right)$ and $R^*(a) = \omega^{-1}\left(\omega(u(a)) \cdot e^{R(a)}\right) - u(a)$.

Finally, if $R(a) > 0$ in the previous argument, then $\omega(u(a) + R(a)) > \omega(u(a))$ and hence $R^*(a) > 0$. An analogous argument holds for the converse.

A.5. *Proof of Corollary 3*

The argument is analogous to the proof of Theorem 1, with the strictly increasing cdf G taking the place of the function $1/(1 + e^{-x})$.

APPENDIX B: Independence of Logit-Reinforcement Axioms

In view of Theorem 1 and Corollary 1, axiom BCA is obviously independent of SDA, IDA, and HI. Any logit-reinforcement model where $R(a)$ is not always positive will fulfill SDA, IDA, and HI, but fail BCA (for instance, Example 4). The following examples show that axioms SDA, IDA, and HI are independent of each other, even when BCA is satisfied.

EXAMPLE 9: Axioms SDA, HI, and BCA do not imply IDA. Consider an SCM-U where the SCM captures uncertainty through utility functions, $\Theta = \mathcal{U}$, and choice probabilities $P_u(a,b)$ are determined through the logit choice function (2) for each u. However, updating does not follow reinforcement as considered above. Instead, let $R^* : \mathcal{A} \mapsto \mathbb{R}_{++}$ be a function assigning a positive value to the chosen option for each *choice*, $R(a,b) > 0$. For each $u \in \mathcal{U}$, let $f(u,\emptyset) = u$. For each $(a,b) \in \mathcal{A}$, let $f(u,(a,b)) = u_{a,b}$ be given by

$$u_{a,b}(c) = \begin{cases} u(a) + R(a,b) & \text{if } c = a \\ u(c) & \text{if } c \neq a. \end{cases} \tag{9}$$

for each $c \in A$.

Given some prior utility u, consider the DSCF generated by this SCM-U. SDA is obviously satisfied because the utilities of options other than the chosen one are not updated. HI also holds, because the update in utilities does not depend on the previous history. BCA follows because $R(a,b) > 0$. However, IDA fails in general. For example, if $A = \{a,b,c,d\}$ with $u(x) = 0$ for all $x \in A$, $R(a,b) = \ln 2$ and $R(a,c) = \ln 3$, then $P(a,d|a,b) = 2/3 \neq 3/4 = P(a,d|a,c)$.

EXAMPLE 10: Axioms IDA, HI, and BCA do not imply SDA. As in the last example, consider an SCM-U with $\Theta = \mathcal{U}$ such that choice probabilities $P_u(a,b)$ are determined through the logit choice function (2) for each u. Updating occurs as follows. For each $u \in \mathcal{U}$, let $f(u,\emptyset) = u$, and define $M_u = 1 + \max\{u(x) - u(y) \mid (x,y) \in \mathcal{A}\}$. By symmetry of $\mathcal{A}$, $M_u > 0$. For each $(a,b) \in \mathcal{A}$, let $f(u,(a,b)) = u_{a,b}$ be given by

$$u_{a,b}(c) = \begin{cases} 2 \cdot u(a) + M_u & \text{if } c = a \\ 2 \cdot u(c) & \text{if } c \neq a. \end{cases} \tag{10}$$

for each $c \in A$. Given some prior utility u, consider the DSCF generated by this SCM-U. IDA and HI are satisfied because the update in utilities does not depend on the identity of the rejected option or on the previous history. To see BCA, let $(a,b) \in \mathcal{A}$. For the history $h \in \mathcal{H}$, let $u^h = f(u,h)$. Then, under logit choice, $P_{f(u,(a,b)\circ h)}(a,b) > P_{f(u,h)}(a,b)$ holds if and only if

$$2\left(u^h(b) - u^h(a)\right) + M_{u^h} > u^h(b) - u^h(a),$$

or, equivalently, $M_{u^h} > u^h(a) - u^h(b)$. The latter holds by construction. However, this DSCF does not satisfy SDA. Let, for instance, $A = \{a,b,c\}$ and $u(b) = \ln 2$, $u(c) = 1$. Then, $P(b,c|a,b) = 4/5 \neq 2/3 = P(b,c|\emptyset)$.

EXAMPLE 11: Axioms SDA, IDA, and BCA do not imply HI. Consider again the set of model parameters $\mathcal{U}$ and assume that choice probabilities $P_u(a,b)$ are determined through the logit choice function (2) for each u. Consider a history-dependent reinforcement function $R^* : \mathcal{A} \times \mathcal{H} \mapsto \mathbb{R}_{++}$ such that $R^*((a,b),h) = \ell(h) + 1$, where $\ell(h)$ denotes the length of h. That is, for each $u \in \Theta$, $(a,b) \in \mathcal{A}$ and $h \in \mathcal{H}$, let $u_{a,b,h}$ denote the updated utility given by

$$u_{a,b,h}(c) = \begin{cases} u(a) + \ell(h) + 1 & \text{if } c = a \\ u(c) & \text{if } c \neq a. \end{cases} \tag{11}$$

for each $c \in A$. Hence, utilities are updated as in the logit-reinforcement model, but the reinforcers depend on the entire history. Given any prior utility u, iteration of this procedure

generates a DSCF.[11] BCA follows, because $R^*(a,b) > 0$ for all (a,b). SDA and IDA follow because they only involve comparisons for a fixed history. However, HI fails in general because the updated utilities depend on the length of the history. For instance, suppose $a, b \in \mathcal{A}$ and $u_\emptyset(a) = u_\emptyset(b) = 0$. Then $P(a,b) = 1/2$, $P(a,b|a,b) = e/(1+e)$, and $P(a,b|((a,b),(a,b))) = e^3/(1+e^3)$. Thus, taking $h = \emptyset$, the odds quotient on the left-hand-side of (5) (with $c = b$) is equal to e, but taking $h' = (a,b)$, the quotient on the right-hand-side is equal to e^2. Thus, the DSCF does not satisfy HI.

APPENDIX C: Updated choice probabilities for Example 7

For Example 7, the following table provides the choice probabilities after a single round of updating based on any possible choice observation.

Updated probabilities	Observation (a,b') or (a,b')	(b,a) or (b',a)	(a,c) (a,c)	(c,a) (c,a)	(b,b') (b,b')	(b',b) (b',b)	(b,c) or (b',c)	(c,b) or (c,b')
(a,b) or (a,b')	1	0	2/3	1/3	1/2	1/2	1/3	2/3
(a,c)	2/3	1/3	1	0	1/2	1/2	2/3	1/3
(b,b')	1/2	1/2	1/2	1/2	1	0	1/2	1/2
(b,c) or (b',c)	1/3	2/3	2/3	1/3	1/2	1/2	1	0

TABLE C.I

UPDATED PROBABILITIES IN EXAMPLE 7

APPENDIX D: Further Implications of the ABSP

As Theorem 2 shows, the ABSP fully characterizes the RU-Bayesian model. However, since the characterization follows a very different approach to the logit-reinforcement model, direct comparisons with the axioms for that model are difficult. We therefore consider some implications of the ABSP which highlight the commonalities and differences with the logit-reinforcement model.

As shown in Section 5.4, a first immediate implication of the ABSP is that, following any admissible history, stochastic choices must satisfy the ARSP (Axiom U-ARSP). A second, straightforward implications of the ABSP is determinism, as also mentioned in Section 5.4. That is, once a choice (a,b) is observed, the probability of the opposite choice (b,a) is updated to zero, i.e., the choice among a and b becomes deterministic. Formally:

AXIOM—DET: *For all $s \in \bar{\mathcal{A}}$ and all $h \in \bar{\mathcal{H}}$ such that $P(s|h) > 0$, $P(s|s \circ h) = 1$.*

It is straightforward to show that Axiom DET is implied by the ABSP: consider any admissible history h and choice s such that $P(s|h) > 0$. Since h is admissible, it follows by definition that $\Pi(h) > 0$. Moreover, the histories $s \circ h$ and $s \circ s \circ h$ are consistent with exactly the same set of maximally tight histories, i.e., $\mathcal{H}^*(s \circ h) = \mathcal{H}^*(s \circ s \circ h)$. Therefore, the ABSP implies

[11] The procedure to generate a DSCF described in this example is not formulated as a SCM-U. However, Proposition 1 implies that some SCM-U can rationalize the DSCF generated by the procedure for each prior parameter $u_\emptyset \in \mathcal{U}$.

$P(s|s \circ h)P(s|h)\Pi(h) = P(s|h)\Pi(h)$, and hence $P(s|s \circ h) = 1$. However, the DSCF generated from a logit-reinforcement model cannot satisfy Axiom DET as this is inconsistent with positivity (i.e., Axiom U-POS). Axiom DET therefore provides a simple condition to distinguish between the RU-Bayesian and logit-reinforcement models in terms of a single round of updating.

A third implication of the the ABSP is reminiscent of the Luce Product Rule (LPR) but concerns updated probabilities across cycles of three alternatives. We refer to it as the Bayesian Product Rule (BPR).

AXIOM—BPR: *For all distinct $a, b, c \in A$ and all $h \in \bar{\mathcal{H}}$ such that $P(x, y|h) > 0$ for $x, y \in \{a, b, c\}$,*

$$P(a,b|(b,c) \circ h)P(b,c|(c,a) \circ h)P(c,a|(a,b) \circ h) = P(a,b|(c,a) \circ h)P(c,a|(b,c) \circ h)P(b,c|(a,b) \circ h)$$

and

$$P(a,c|(c,b) \circ h)P(c,b|(b,a) \circ h)P(b,a|(a,c) \circ h) P(a,c|(b,a) \circ h)P(b,a|(c,b) \circ h)P(c,b|(a,c) \circ h).$$

The BPR follows immediately from the ABSP by recognizing that, for instance, the histories $(a, b) \circ (b, c) \circ h$ and $(b, c) \circ (a, b) \circ h$ are consistent with the same set of maximally tight histories, i.e., $\mathcal{H}^*((a, b) \circ (b, c) \circ h) = \mathcal{H}^*((b, c) \circ (a, b) \circ h)$. However, the following example shows that the BPR can be violated by a logit-reinforcement model.

EXAMPLE 12: Let $A = \{a, b, c\}$ and consider a utility function given by $u(a) = \ln 3$, $u(b) = \ln 2$, and $u(c) = 0$. Consider the prior probabilities $P(\cdot|\emptyset)$ derived from a logit model with utility u, and the updated probabilities $P(s|s')$ defined by a logit-reinforcement model with reinforcement function given by $R(a) = \ln C - \ln 3$, $R(b) = \ln 3 - \ln 2$, and $R(c) = \ln 2$, for some $C \geq 1$.

	Observation						
Updated probabilities	$\emptyset$	(b,c)	(c,b)	(a,c)	(c,a)	(a,b)	(b,a)
(a,b)	$3/5$	$1/2$	$3/5$	$C/(C+2)$	$3/5$	$C/(C+2)$	$1/2$
(b,c)	$2/3$	$3/4$	$1/2$	$2/3$	$1/2$	$2/3$	$3/4$
(c,a)	$1/4$	$1/4$	$2/5$	$1/(C+1)$	$2/5$	$1/(C+1)$	$1/4$

This defines a collection of one-period-updated SCFs which fulfill property U-ARSP. However, the collection of SCFs does not satisfy the BPR because

$$P(a,b|b,c)P(b,c|c,a)P(c,a|a,b) = \frac{1}{2}\frac{1}{2}\frac{1}{C+1}$$

and

$$P(a,b|c,a)P(c,a|b,c)P(b,c|a,b) = \frac{3}{5}\frac{1}{4}\frac{2}{3} = \frac{1}{10}.$$

Hence, the first equality in BPR holds only for $C = 1.5$. Analogously, the second identity holds only for $C = 1$, and hence the BPR is not satisfied for any C.

The BPR is therefore another property of a DSCF that can be used to distinguish between the RU-Bayesian and logit-reinforcement models based on updating from a single new choice observation. However, axioms U-ARSP, DET and BPR cannot be sufficient to characterize the RU-Bayesian model as, given a history, they pertain only to updating from at most one additional choice observation. The following example illustrates that the axioms taken together are not sufficient to characterize the RU-Bayesian model even for the case of just three alternatives.

EXAMPLE 13: Let $A = \{a, b, c\}$ and consider a collection of SCF as given by the following (updated) choice probabilities, for $\varepsilon \in (0, 1)$.

	Observation						
Updated probabilities	$\emptyset$	(b,c)	(c,b)	(a,c)	(c,a)	(a,b)	(b,a)
(a,b)	$1/2$	1	0	1	0	1	0
(b,c)	$1/2$	1	0	1	0	$1-\varepsilon$	0
(c,a)	$1/2$	1	0	1	0	ε	0

The U-ARSP reduces to the constraint that the updated probabilities of (a,b), (b,c), and (c,a) given any fixed observation (possibly $\emptyset$) add up to at most two, which is true by direct inspection of the table above. Property DET also holds immediately. As for BPR,

$$P(a,b|b,c)P(b,c|c,a)P(c,a|a,b) = 0 = P(a,b|c,a)P(c,a|b,c)P(b,c|a,b)$$

and analogously for the second identity. Hence, all three necessary conditions identified above hold. Suppose this collection of SCFs could be rationalized in terms of a RUM with Bayesian updating. Then, it follows that

$$P(a,b|b,c)P(b,c|\emptyset) = P(b,c|a,b)P(a,b|\emptyset)$$

which, in view of the choice probabilities above, yields the contradiction $\varepsilon = 0$. Hence, the collection of SCFs is not consistent with a DSCF that is rationalized by the RU-Bayesian model.

For any model that a researcher might develop, a question of immediate interest is whether the model can be encompassed by previous ones. Recall, for instance, the random-RU-Bayesian model (Example 8). As discussed in Section 5.4, this model fails the ABSP but fulfills its implication, the U-ARSP. Also, a DSCF rationalized by the random-RU-Bayesian model with prior θ_0 satisfies Axiom DET if and only if the support of θ_0 is concentrated on Dirac distributions over strict preferences. The random-RU-Bayesian model is also not observationally equivalent to a logit reinforcement model. On the one hand, since random-RU-Bayesian models encompass RU-Bayesian models, the DSCFs generated by the model do not in general satisfy the Luce Product Rule (Axiom U-LPR), which is always satisfied by the DSCF generated by a logit-reinforcement model. On the other hand, it is straightforward to derive from Bayes' rule that any DSCF that is rationalized by the random-RU-Bayesian model always satisfies the Bayesian Product Rule (BPR), which can be violated by logit-reinforcement models as shown above (Example 12). The random-RU-Bayesian model therefore provides an example of a stochastic choice model with updating that is in the spirit of the Bayesian perspective on dynamic stochastic choice, but potentially allows for choices to remain stochastic even after being updated.

The following example is a different generalization of the Bayesian approach (that is, different from the random-RU-Bayesian model) that also allows for choice predictions to remain stochastic even with an accumulation of past choice observations, hence violating Axiom DET.

EXAMPLE 14—Trembling-Hand Model: Fix $\varepsilon \in [0,1]$. Let $\Theta = \Delta(\mathcal{R})$, where, as in the RU-Bayesian model, $\mathcal{R}$ represents the set of strict preferences over the set of alternatives A. For $\succ \in \mathcal{R}$, define

$$P^{\varepsilon}_{\succ}(a,b) = \begin{cases} 1-\varepsilon & \text{if } a \succ b \\ \varepsilon & \text{if } b \succ a \end{cases}.$$

Given $\pi \in \Delta(\mathcal{R})$, define $P^{\varepsilon}_{\pi}(a,b) = \sum_{\succ \in \mathcal{R}} \pi(\succ) P^{\varepsilon}_{\succ}(a,b)$. The collection $\{P^{\varepsilon}_{\pi}\}_{\pi \in \Delta(\mathcal{R})}$ defines a SCM which can be interpreted as follows. An external observer holds beliefs over the actual preferences of a decision maker, $\succ$. However, the observer assumes that the decision maker has a "trembling hand" and might always make decisions against the actual preference, with probability ε.

Suppose the external observer is Bayesian, but respects the structural assumptions of the model, including the error probability ε.[12] When confronted with a new choice (a,b), the observer must consider the possibility that the choice has been a mistake. Hence, the updating function $f^{\varepsilon}_{B} : \Delta(\mathcal{R}) \times \mathcal{A}^{*} \to \Delta(\mathcal{R})$ is given by

$$f^{\varepsilon}_{B}(\pi, s) = \begin{cases} \pi & \text{if } s = \emptyset \\ \pi^{\varepsilon}_{s} & \text{if } s = (a,b) \in \mathcal{A}, \end{cases}$$

where, for each $(a,b) \in \mathcal{A}$, $\pi^{\varepsilon}_{a,b}(\succ) = \frac{\pi(\succ) \cdot P^{\varepsilon}_{\succ}(a,b)}{P^{\varepsilon}_{\pi}(a,b)}$. Let $P_{\pi}(a,b)$ be the error-free probability as given in Example 3. Note that, if $s = (a,b) \in \mathcal{A}$ with $P_{\pi}(a,b) = 0$, then $P^{\varepsilon}_{\succ}(a,b) = \varepsilon$ for all $\succ \in \mathcal{R}$ with $\pi(\succ) > 0$, and it follows that $f^{\varepsilon}_{B}(\pi, s) = \pi$. That is, if the observed choice contradicts the prior, the observer concludes that it was a mistake and does not update.

The tuple $(\Delta(\mathcal{R}), \{P^{\varepsilon}_{\succ}\}_{\succ \in \mathcal{R}}, f^{\varepsilon}_{B})$ is a stochastic choice model with updating, which we call the *trembling-hand model*.

When $\varepsilon \in \{0,1\}$, the Trembling-Hand model reduces to the RU-Bayesian model, but not when the error probability $\varepsilon \in (0,1)$, because the DSCF generated by a trembling-hand model then clearly violates Axiom DET. Indeed, the DSCF generated by a trembling-hand model satisfies Axiom DET if and only if $\varepsilon \in \{0,1\}$ and, otherwise, satisfies Axiom U-POS. As such, strict trembling-hand models, with an error probability $\varepsilon \in (0,1)$, retain the basic intuitions of the Bayesian approach but satisfy the positivity condition of logit-reinforcement models. However, the strict trembling hand model is not encompassed by the logit-reinforcement model either because the stochastic choices with updating do not, in general, satisfy the Luce Product Rule U-LPR. For instance, suppose $A = \{a,b,c\}$ and π places probability one on $a \succ b \succ c$. The probability of the intransitive cycle $a \to b \to c \to a$ is $P(a,b)P(b,c)P(c,a) = (1-\varepsilon)^2\varepsilon$, while the probability of observing the cycle $a \to c \to b \to a$ is $P(a,c)P(c,b)P(b,a) = (1-\varepsilon)\varepsilon^2$. Hence, in view of Lemma 1, the strict trembling hand model differs from the logit-reinforcement model in terms of the prior probability of choices, even without considering updating.

[12] A more general model could incorporate ε into Θ, allowing the external observer to update its value. For concreteness, we consider ε to be exogenous.