Submitted to Econometrica

Dynamic constraints on the distribution of stochastic choice

Ryan Webb

March 12, 2015

Rotman School of Management, University of Toronto, 105 St George St., Toronto, Ontario, M6G2M5. E-mail: ryan.webb@utoronto.ca

Submitted to Econometrica

DYNAMIC CONSTRAINTS ON THE DISTRIBUTION OF STOCHASTIC CHOICE

Ryan Webb[1]

Keywords: Neuroeconomics, Stochastic Choice, Random Utility, Bounded Accumulation, Drift Diffusion, Discrete Choice Econometrics; JEL Classification: D87, C12.

PLEASE DO NOT CIRCULATE

The *random utility* model is the standard empirical framework for modelling stochastic choice behaviour in applied settings. In this framework, the distribution of random utility has important implications both for testing behavioural theories and predicting behaviour, however the theoretical and empirical foundations of this distribution are not well understood. Moreover, the random utility framework has so far been agnostic about the dynamics of the decision process that are of considerable interest in psychology and neuroscience, captured by a class of *bounded accumulation* models which relate decision times to stochastic behaviour. We demonstrate that the random utility model can be derived from a bounded accumulation model, of which the classic *drift diffusion* model is a special case. The features of this dynamic process impact the distributional structure of random utility and potentially bias estimates of structural choice parameters. Finally, the article demonstrates how particular features of the bounded accumulation frameworks can constrain the distribution of random utility, offering advances in modelling choice behaviour.

1. INTRODUCTION

Stochastic choice behaviour is an established empirical phenomenon and the Random Utility Model (RUM; Marschak, 1960; Becker, DeGroot, and Marschak, 1963) has become the standard framework for modelling it in applied settings (McFadden, 2001). The RUM has proven successful because

[1]The author wishes to thank P.W. Glimcher, R. Kiani, T. LoFaro, K. Menzel, A.A.J. Marley, and A. Rangel for helpful comments.

it provides a highly-flexible empirical framework for relating observables in a dataset to choice prediction.[1] However, it has been well-documented that the distribution of random utilities has important implications for both predicting choice behaviour and testing behavioural theories, often overshadowing the underlying theory itself (Manski, 1977; Hausman and Wise, 1978; Loomes and Sugden, 1995; McFadden, 2001; Loomes, 2005; Hey, 2005; Wilcox, 2008; Haile, Hortaçsu, and Kosenok, 2008; Train, 2009). The crucial role played by the stochastic element has therefore led to calls for an empirical foundation for stochastic choice models (Loomes, 2005; Hey, 2005), and speculation that neuroeconomics can offer a richer, data-driven, approach (Harrison, 2008).

Over the past fifteen years, a significant amount of progress has been made in understanding the neural processes underlying choice (for reviews, see Glimcher, 2011; Fehr and Rangel, 2011; Glimcher and Fehr, 2013). Particular emphasis has been placed on the dynamics of decision-making since the timing of a decision gives insight into the underlying neural circuitry (Roitman and Shadlen, 2002; Wang, 2002; Gold and Shadlen, 2007) and improves the prediction of choice behaviour (Milosavljevic, Malmaud, Huth, Koch, and Rangel, 2010; Krajbich, Armel, and Rangel, 2010). To address this relationship between time and choice, a class of Bounded Accumulation Model (BAM) – drawn from psychophysical and neuroscientific work on sensory decision-making – has received much attention (for reviews see Ratcliff and Smith, 2004; Gold and Shadlen, 2007). Of this large model class, the Drift Diffusion Model (DDM; Ratcliff, 1978) is the most well-known special

[1] A RUM is an implementation of stochastic revealed preference, a population analog of revealed preference theory which seeks to identify whether distributions of observed choices are consistent with a hypothesis of utility maximization within a population (McFadden, 2005). Here, we return to the original conceptualization of a RUM by explicitly interpreting the population as consisting of multiple choices by the same individual in the same choice situation (McFadden, 1981).

case.

The goal of this article is to demonstrate that the general class of BAM implies a form of RUM, to examine the restrictions the BAM imposes on the distribution of stochastic choice, and to provide methods for incorporating these results into empirical application. After briefly reviewing the two model classes in Section 2, our formal statement linking them is provided in Section 3. The intuition is straight-forward: a BAM implements the $\max\{\}$ operator in a utility maximization, but with some element of stochasticity. Moreover, the implications that a particular BAM has for the distribution of stochastic choice depend crucially on its specification. In particular, the distribution of stochastic choice is influenced by the dynamics of the choice process, a subject on which the discrete choice literature has so far remained agnostic.[2]

To clarify this point, in Section 4 we pursue a number of examples which vary in their degree of analytical tractability. In the case of the DDM, it is well-known that closed-form solutions for the choice probabilities reduce to the familiar Logit formulation (Cox and Miller, 1965). Our result for the DDM demonstrates that the Logit can be derived from a larger class of error distributions than typically stated in the discrete choice literature (McFadden, 1974). In the case of a "fixed-threshold" BAM, closed forms exist for the distributions of key random variables. Our results demonstrate that this class of BAM yields a RUM in which the variance of stochastic choice depends on observables. For the general class of BAM, neither the choice probabilities nor the distribution of key random variables can be characterized in closed form. In this case, we present a method for numerically approximating choice probabilities (and the distributions of other potentially observable data such as decision time). How well this approxi-

[2]More accurately, standard models make distributional assumptions which ignore the role of dynamics explored in this article.

mation captures the actual choice process – and choice data in particular – then becomes an empirical question.

Notably, the observation of decision time provides additional information which improves the efficiency, or reduces the bias, of model estimates. In section 5, we pursue an example which demonstrates the econometric implications of our results. If choices are generated using a BAM, estimates of common structural parameters, such as the coefficient of relative risk aversion, can be biased in a well-known experimental condition for choice under uncertainty (Holt and Laury, 2002). This bias can be partially corrected using established techniques. Incorporating an observation of decision time into the stochastic specification reduces the bias, while estimating a specification derived directly from the BAM eliminates the bias.

The incorporation of bounded accumulation models into the random utility framework therefore has as important implications for discrete choice modelling. For economists, it provides a means of constraining a stochastic choice model to the empirical features of the choice process, thereby guiding applied research. In the other direction, the abstraction provided by the RUM offers a number of econometric advantages that would be impractical through modelling at the level of neural dynamics alone (Webb, Glimcher, Levy, Lazzaro, and Rutledge, 2013). This interaction emphasizes the gains from modelling choice behaviour at different levels of abstraction.

2. MODELS OF STOCHASTIC CHOICE

2.1. *The Random Utility Model*

Consider a choice set comprised of n alternatives (indexed $i = 1 \dots n$) and a vector of measured observables $\mathbf{x}_i$ for each alternative. Formally, the RUM posits a vector of random variables $\mathbf{u}$, with element u_i, such that $\Pr[u_{i^*} > u_j, \quad \forall j \neq i^*]$ equals the probability that alternative i^* is chosen from the choice set; equivalently stated via the choice criterion,

$$i^* = \arg\max_i \{u_i\} .$$

In this general formulation, conditions placed on the choice probabilities determine whether observed behaviour is consistent with the principle of utility maximization (Block and Marschak, 1960; McFadden, 1974; Falmagne, 1978). The formulation of a RUM as an empirical tool was made possible by applying an insight from psychophysics - the probability of choosing an alternative depends on the discriminability (or magnitude difference) between stimuli (Weber, 1834; Fechner, 1860; Thurstone, 1927) - to utility in a consumer choice problem (Becker, DeGroot, and Marschak, 1963; McFadden, 1981). Formally, this yields an additive form for $\mathbf{u}$, composed of deterministic valuations $\mathbf{v}$ determined by the theory under investigation, and a random vector $\boldsymbol{\eta}$ defined independently of $\mathbf{v}$. The role of the behavioural theory is now explicitly stated via $v_i = f(\mathbf{x}_i, \beta_i)$, and the choice criterion is given by:

$$\begin{aligned} (1) \qquad i^* &= \arg\max_i \{v_i + \eta_i\} \\ &= \arg\max_i \{f(\mathbf{x}_i, \boldsymbol{\beta}) + \eta_i\} . \end{aligned}$$

This additivity assumption yields the "Fechner" model that has found application throughout the discrete choice literature (McFadden, 2001).[3] Its usefulness arises from a convenient specification of the choice probabilities: the probability of choosing an alternative i depends on the comparison between the magnitude $v_i - v_j$ and the random variable $\tilde{\eta}_{ji} \equiv \eta_j - \eta_i$,

$$(2) \qquad P_i(\mathbf{v}) = \Pr\left[v_i - v_j > \tilde{\eta}_{ji}, \quad \forall j \neq i\right] .$$

[3]In a Neural Random Utility Model (Webb, Glimcher, Levy, Lazzaro, and Rutledge, 2013), v_i is no longer an unobservable latent variable. Instead it is an observable neural quantity, and the model is formulated directly in terms of these observations as in equation (1). This eliminates sources of error due to the econometrician's inability to directly observe valuations in a standard dataset (Manski, 1977), and can lead to improved choice prediction provided suitable measurement techniques.

Only an assumption for the distribution of $\boldsymbol{\eta}$ is required. For instance, if η_i is assumed to be Gaussian and independent over alternatives, then $\tilde{\eta}_{ji}$ is Gaussian and the choice probabilities are given by the Probit model. If η_i follows an independent Gumbel distribution, $\tilde{\eta}_{ji}$ is distributed logistic and the choice probability are given by the logistic function underlying the Multinomial Logit model (sufficiency in Luce and Suppes, 1965; McFadden, 1974, for necessity).

Since the additive formulation of (1) implies that a $\mathbf{u}$ can be defined for any $\mathbf{v}+\boldsymbol{\eta}$, every Fechner model is a RUM (Becker, DeGroot, and Marschak, 1963; Batley, 2008). However, the reverse implication does not hold. An example of this which nicely demonstrates the importance of correctly specifying the distribution of $\boldsymbol{\eta}$ can be found in the experimental literature on choice under uncertainty. For example, one might assign $v_i = EU_i(\mathbf{x}_i, \mathbf{p}_i; \alpha)$ to be the expected utility of a lottery $(\mathbf{x}_i, \mathbf{p}_i)$ parametrized with a risk preference α. Under this interpretation, the choice probabilities of a Fechner RUM (2) are determined by the magnitude difference of the valuations that EU places on each lottery. Therefore choice behaviour deviates from EU in accordance with the covariance structure placed on $\boldsymbol{\eta}$. For the purposes of testing theory, the Fechner RUM is, essentially, a model of errors.

This class of RUM lies in contrast (or in addition to) an alternative approach which has a rich history in the theoretical and experimental literature. Loosely speaking, instead of a single utility function that is compared with some error, the decision-maker has multiple possible utility functions (which obey some axiom(s) under question) and one function is drawn on each choice trial with some probability. In the experimental and decision theory literature, this approach is referred to as a Random Preference (RP) model (Loomes and Sugden, 1995, 1998).[4] Returning to our example, the

[4]In the theoretical literature, this approach is also termed a *random utility* model (e.g. Gul and Pesendorfer, 2006), since the probability measure is over a set of utility functions.

risk preference α would be a stochastic quantity, and the comparison between (the now random) v_i would be made without error on each trial.[5]

The distinction between a Fechner RUM and a RP model has important behavioural predictions. With regard to testing EU, a RP model would predict no violations of First Order Stochastic Dominance (FOSD) since the utility function is drawn from the set of expected utility functions. In contrast, a Fechner RUM would predict some FOSD violations depending on the degree of covariance of $\boldsymbol{\eta}$. Attempts to resolve this debate in the experimental literature have led to a mixed conclusion. Though FOSD violations are observed in choice experiments – (strictly) ruling out the RP model – the pattern of violations can not be reconciled by a Fechner RUM where the η_i are independent, have constant variance, and/or are symmetric (Loomes, 2005). Indeed, the variance of $\tilde{\eta}_{ji}$ seems to depend on the time a decision is made (Hey, 1995) as well as other observable features of the set of lotteries (Buschena and Zilberman, 2000).

This one example typifies a larger issue in the discrete choice literature, namely that the stochastic specification has important implications for testing theory. In a thorough analysis of the interaction between stochastic choice and the structural theories of choice under risk and uncertainty, Wilcox (2008) concludes that the stochastic element of choice is "at least as important in determining sample properties as structures are." This general point is made clear by Loomes and Sugden (1995), who note that "alternative ways of modelling the stochastic element are not neutral with respect to the hypotheses they entail, and future theoretical and empirical work should not regard the stochastic specification as an 'optional add-on', but

This naming convention often causes confusion when moving between the theoretical and empirical literatures.

[5]Of course, repeated choices over the same lotteries would still differ due to the stochasticity in α. For an in-depth discussion of the interpretation of Random Preference models in terms of neural variables, see Webb, Glimcher, Levy, Lazzaro, and Rutledge (2013).

rather as an integral part of every theory which seeks to make predictions about decision-making under risk and uncertainty." Such renewed focus on the stochastic element of choice presents an opportunity for insight from neuroeconomics, with structural error parameters and more flexible specifications that are driven by both neural and behavioural data (Harrison, 2008).

The goal of this article is to do just this: to provide a foundation for the distribution of stochastic choice derived from the current neuroscientific approaches to modelling the dynamics of decision-making. We demonstrate that a Fechner RUM can be derived from a general class of Bounded Accumulation Models, and explore how various special cases impact the distribution of η_i and the dependence of its moments on $\mathbf{v}$. [6]

2.2. *Bounded Accumulation Models*

The dynamic process by which the brain reaches a decision is a topic of intense research interest in psychology and neuroscience. Originally, this work focused on the dynamics of *perceptual* decision-making in which subjects are required to determine the state of an objectively known stimulus (e.g. the direction of motion of erratically moving dots; Britten, Shadlen, Newsome, and Movshon, 1992). The class of Bounded Accumulation Models (BAMs) have come to address how such sensory evidence is accumulated over time by neural activity, integrated into a decision signal, and how a decision is implemented once this signal reaches some boundary (for reviews see Ratcliff and Smith, 2004; Smith and Ratcliff, 2004; Gold and Shadlen, 2007).

[6] However we should note that we do not exhaust the set of neuroscientific models that fall in the domain of a RUM. For instance, Webb, Glimcher, and Louie (2014) explores how neurobiological constraints on the processing of information, or equivalently, the bounding of the utility representation, implies observed choice probabilities which violate the IIA axiom.

Recently, BAMs have been extended to "value-based" or "goal-directed" choice environments familiar to economists, in which subjective valuations of common consumer goods replace sensory input (Milosavljevic, Malmaud, Huth, Koch, and Rangel, 2010; Krajbich, Armel, and Rangel, 2010; Fehr and Rangel, 2011; Webb and Dorris, 2013).

We now layout the general formulation for a BAM that is composed of three elements: a stochastic process which accumulates subjective value, a stopping rule, and a choice criterion. The class of such models is large, with subsequent incarnations able to better capture both behaviour and the dynamics of neural activity. Much of the focus has been on grounding the dynamic process in principles of neural computation (e.g. leakage and mutual inhibition, Usher and McClelland, 2001; Wang, 2002; Ditterich, Mazurek, Roitman, and Shadlen, 2003; Wong and Wang, 2006; Roxin and Ledberg, 2008; Bogacz, Brown, Moehlis, and Holmes, 2006, for an overview). However there is still much ongoing debate over the particulars of the dynamic process, the form of the boundary, and how a dynamic decision process can be implemented in neural architecture (Cisek, 2006; Kiani, Hanks, and Shadlen, 2008; Tsetsos, Usher, and Chater, 2010; Ditterich, 2010; Hunt, Kolling, Soltani, Woolrich, Rushworth, and Behrens, 2012; Tsetsos, Gao, McClelland, and Usher, 2012; Ditterich and Churchland, 2012; Heitz and Schall, 2012; Liston and Stone, 2013; Kiani, Corthell, and Shadlen, 2014).

Our formal statement of the class of BAMs is designed to capture the literature as generally as possible, in a manner that is tractable to relate to a RUM. In doing so, we will note special cases which correspond to well-known models in the neuroscience and psychology literature. Each of these special cases will be comprised of a particular accumulation process and stopping rule. In some cases these models will have familiar, closed-form, solutions for the accumulation processes and/or expressions for the choice probabilities. In the remaining cases, numerical approximations exist.

2.2.1. *The Accumulation Process*

Each alternative i in the choice set is associated with a decision signal $Z_i(t)$ which accumulates v_i over time. Together, the $Z_i(t)$ form the $n \times 1$ vector $\mathbf{Z}(t)$. In empirical practise, this signal is taken to be the activity level of a population of neurons associated with alternative i. In a formal model, $\mathbf{Z}(t)$ is a vector-valued Markov Process in the continuous state space $\mathbb{R}^n$, for a continuous time index $t \in T$.[7] While the continuous time process can be derived as the limit of a discrete time random-walk (Shadlen, Hanks, Churchland, and Kiani, 2006), the continuous time version has the advantage that it includes the class of Gaussian processes for which solutions can be derived in closed form.[8]

For brevity, we skip straight to a formulation of the accumulation process in terms of stochastic differential equations.[9]

Definition 1 *A general accumulation process $\mathbf{Z}(t)$ accumulates $\mathbf{v}$ according to*

$$d\mathbf{Z}(t) = [\mathbf{v} + \nu(t) + \mathbf{\Gamma}(t)\mathbf{Z}(t)]\, dt + \boldsymbol{\sigma}(t)\, d\mathbf{B}(t), \tag{3}$$

where $\mathbf{B}(t)$ is a (vector-valued) Gaussian process with independent increments (i.e. Brownian motion, Mörters and Peres, 2010). The diffusion of $\mathbf{B}(t)$ is governed by the ($n \times n$ matrix-valued) function $\boldsymbol{\sigma}(t)$, allowing for

[7] Given the interpretation of $\mathbf{v}$ as an observable neural quantity, it is natural to assume $v_i > 0$. We also assume that $Z_i(0) = 0$ and that $Z_i(t)$ can take on negative values. A BAM primarily concerned with neural implementation impose the constraint that $Z_i(t) > 0$ by setting positive baseline activity $Z_i(0) > 0$ and a reflective bound at zero.

[8] Discrete time may, in fact, be the more appropriate approach to modelling neural activity (Shadlen, Hanks, Churchland, and Kiani, 2006). Our main result can be shown for a discrete state space and discrete time (see working paper, (Webb, 2013)).

[9] We refer readers interested in the technical details of constructing the differential equations, and their solutions, to Cox and Miller (1965)Feller, KarlinTaylor, and in particular, Smith (2000).

time-varying correlation between the stochastic process for each alternative. Similarly, the ($n \times n$ matrix-valued) function $\mathbf{\Gamma}(t)$ allows for both a time- and state-varying relationship in the accumulators. This general formulation also includes a time-dependent drift term $\nu_i(t) = \nu(t) > 0$, $\forall i$, (often termed an "urgency" signal) which, importantly, does not depend on $\mathbf{v}$.

We impose one additional assumption on the stochastic process, namely that the parameters that govern the statistics of the process and the accumulation are equivalent for each alternative.

Assumption 1 *The symmetric $n \times n$ matrices $\mathbf{\Gamma}(t)$ and $\boldsymbol{\sigma}(t)$ can be written in the form $a(t)\mathbf{1} + (b(t) - a(t))\mathbf{I}$, for some $a(t)$ and $b(t)$, where $\mathbf{I}$ is the $n \times n$ identity matrix and $\mathbf{1}$ is a $n \times n$ matrix of ones.*

This assumption is necessary for our results and will be maintained throughout. Of special interest is the case of binary choice with two accumulators governed by $\mathbf{\Gamma} = \left(\begin{smallmatrix} \gamma & \psi \\ \psi & \gamma \end{smallmatrix}\right)$, $\gamma < 0$, $\psi < 0$. Under this process, the signal for each alternative inhibits the other alternative, since $\psi < 0$.

To clarify the general parametrization given in equation (3), it is instructive to consider special cases defined in terms of (sequential) restrictions of the stochastic differential equations. Unlike (3), the solution to each of these cases yields a Gaussian process.

Definition 2 *An accumulation process $Z_i(t)$ is uncoupled if $\mathbf{\Gamma}(t)$ is a diagonal matrix with element $\gamma(t)$.*

This restriction implies that each stochastic process no longer depends on the processes governing the other alternatives. This yields an accumulation process given by n uncoupled differential equations,

$$dZ_i(t) = [v_i + \nu(t) + \gamma(t)Z_i(t)]\,dt + \sigma(t)\,dB_i(t), \quad \forall i, \forall t \geq 0. \tag{4}$$

In neuroscientific terms, the un-coupling of the differential equations precludes *mutual inhibition* between the accumulators. However the statistics of the accumulation are still allowed to depend on both time and state. This is an important requirement. Woodford (2014) demonstrates that optimal information accumulation given a capacity constraint yields an uncoupled process in which the optimal drift depends on the state of the accumulation, as in (4).

Moreover, mush focus has been placed on whether the accumulation perfectly integrates v_i (e.g. Usher and McClelland, 2001). For instance, when $\gamma(t) < 0$ the accumulation is "leaky" in the sense that early realizations of the accumulation are discounted. Eliminating this possibility brings us to a familiar process which is both time- and state-homogenous.

DEFINITION 3 *An accumulation process $Z_i(t)$ is a Brownian motion (with drift) if it is uncoupled, time-homogenous, and $\gamma = 0$.*

$$dZ_i(t) = v_i\, dt + \sigma\, dB_i(t), \quad \forall i, \forall t \geq 0. \tag{5}$$

Brownian motion describes the continuous time properties of a discrete random walk (with finite variance). In particular, it is a Gaussian process with a closed-form solution for $Z(t)$, and can be fully characterized by the mean and covariance of the underlying discrete process(es) (Mörters and Peres, 2010). The mathematical tractability of Brownian motion greatly simplifies the derivation of choice probabilities and other features of the accumulation process. For this reason, it is often applied in practice.

2.2.2. *Stopping Rules*

The second feature of a BAM is a stopping rule which determines when/if each accumulator (or a function of each accumulator) has reached a decision threshold. As with the accumulation process, many forms of stopping rules have been used in the literature. Here, we present the stopping rule in a general form which allows the choice criterion to be generally stated, and present two example rules which have received particular attention.

In the simplest form of a stopping rule, a fixed decision time $\bar{t}$ is imposed and the alternative with the largest $Z_i(\bar{t})$ is chosen.[10] While a fixed stopping rule has the feature of accumulating each v_i over time via a stochastic process, thus introducing stochasticity in the choice, it does not yield a skewed distribution of decision times that is typically observed in the empirical literature (Luce, 1986).

To address this fact, accumulation models implement a rule which terminates the decision once the accumulator reaches some threshold (or equivalently, exits some region). Formally, the stopping time is determined via

$$t^* = \inf\{t : \mathbf{Z}(t) \notin \mathcal{R}(t)\}, \tag{6}$$

for some region $\mathcal{R}(t)$ defined in $\mathbb{R}^n$. The choice i^* is then determined by the largest accumulator at t^*,

$$i^* = \underset{i}{\operatorname{argmax}} \left\{ Z_i(t^*) \right\}. \tag{7}$$

The various accumulation models proposed in the literature differ in how the region $\mathcal{R}(t)$ is specified and yield different properties for the reaction time distribution, primarily how it depends on $\mathbf{v}$ and the decision outcome (i.e. whether the choice was "correct" in identifying the alternative with the largest v_i).

Definition 4 *A Differenced stopping rule is given by equation (6) and a region*

$$\mathcal{R}(t) = \{x_i, x_j : |x_i - x_j| < \theta(t)\}, \quad \forall i \neq j. \tag{8}$$

[10] Experimental conditions which implement this stopping rule are referred to "interrogation protocols" since the subject is not free to determine the decision time.

The Differenced stopping rule terminates the decision when one accumulator is larger than all others by $\theta(t)$.[11] This rule is particularly noteworthy for its application in binary choice when the accumulators are given by Brownian motion and $\theta(t)$ is constant in time.

REMARK 1 *For n=2, if $Z_i(t)$ is a Brownian motion, then $Z_1(t) - Z_2(t)$ is a one-dimensional Brownian motion to dual thresholds.*

A formal proof can be found in Bogacz, Brown, Moehlis, and Holmes (2006); for intuition we can appeal to Figure 1.[12] The differenced stopping rule projects $\mathbf{Z}(t)$ onto the 1-dimensional coordinate axis, $Z_1(t) - Z_2(t)$, and terminates when it reaches either θ or $-\theta$. This model is commonly referred to as a *Drift Diffusion Model* and has been largely successful in matching the properties of reaction time data in binary choice (Ratcliff, 1978; Ratcliff and Smith, 2004).

By definition, the DDM requires that the difference between accumulators is constant at t^*, but puts no restriction on the magnitude of each accumulator. The following stopping rule has the opposite feature: the threshold is fixed, but there is no constraint on the relative magnitudes.

DEFINITION 5 *A Race stopping rule is given by equation (6) and a region*

$$\mathcal{R} = \{x_i : -\infty \le b_i < a_i < \infty\} \quad \forall i. \tag{9}$$

The Race stopping rule traces out a hypercube $\mathbb{R}^n$ that is fixed in time (Figure 1). The simplest interpretation of this rule occurs when $b_i = -\infty$

[11] Technically, the use of the difference operator in the definition of $\mathcal{R}$ projects $Z(t)$ onto the planes spanned by the Z_i and Z_j coordinates. For choice sets larger then two, the threshold region(s) would then be stated using these new coordinates to formally define $\mathcal{R}$.

[12] The discrete random walk case was examined by Link and Heath (1975).

and $a_i = \theta_i$, resulting in a decision when an accumulator surpasses a decision threshold θ_i.

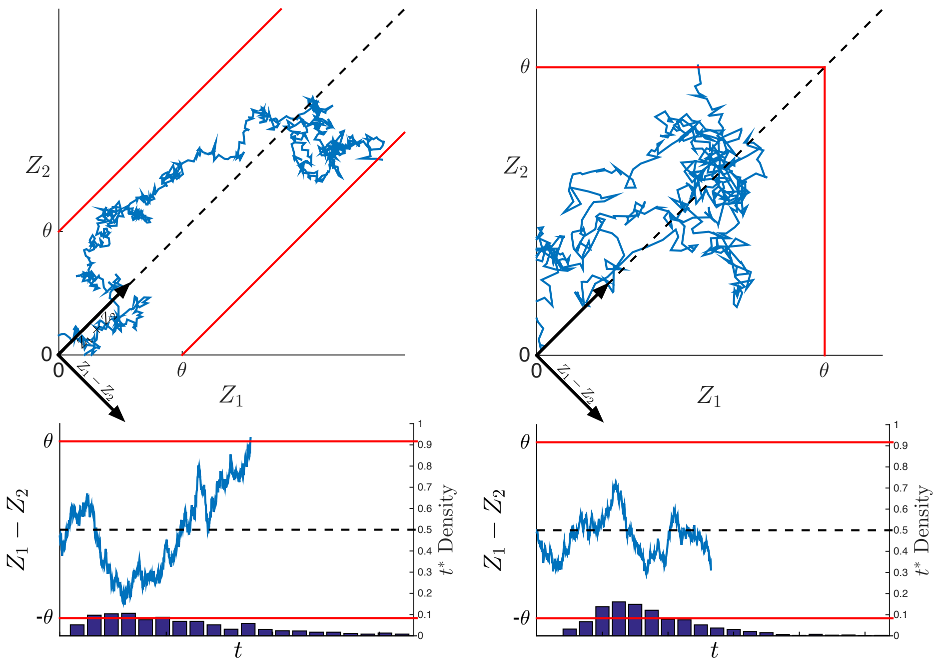

FIGURE 1.— Sample paths of a Brownian motion accumulator for the Differenced (left) and Race (right) stopping rules. Paths of $Z_1(t) - Z_2(t)$ are also depicted as a function of time (below), with the associated density of stopping times.

2.2.3. *State of BAM Literature*

The pairing of an accumulator with a stopping rule defines a particular BAM. We have made an effort to define a general accumulation process (3) and a general stopping rule (6) so that our results will apply as widely as possible to the models explored in the neuroscience and psychology literature. However before we proceed, it will be useful to explore the state of this literature in more detail.

Much of the empirical work in psychology has been focused on capturing the various properties of the decision time distribution. An important feature of all bounded accumulation models is that the accumulation of the decision signal to boundary predicts a positively skewed distribution for decision times (Figure 1) which matches empirical observations (Ratcliff, 1978; Roitman and Shadlen, 2002). In the case of binary choice, the DDM (with the added feature of a stochastic initial condition) is considered the benchmark model (for reviews, see Luce, 1986; Smith and Ratcliff, 2004). The Differenced stopping rule is particularly appealing in the case of a binary choice and no opportunity cost of time. In this case, the DDM implements a normative solution in the sense that it achieves a decision in the least amount of time for a given error rate (Gold and Shadlen, 2002).[13] However when a cost of time is introduced, so that the decision-maker must tradeoff between improving accuracy (by accumulating more evidence) versus moving on to a new decision problem, the optimal solution changes (Drugowitsch, Moreno-Bote, Churchland, Shadlen, and Pouget, 2012). If we assume a binary prior (i.e. the magnitude $|v_1 - v_2|$ is known, but not its sign), the optimal stopping rule collapses the boundaries so that $\theta(t)$ is no longer constant in time. Intuitively, the decision maker is willing to decrease accuracy in the current decision to arrive at a new decision problem earlier.

For the case of three (or more) alternatives, or for priors which do not require $|v_i - v_j|$ to be known, the optimal stopping rule(s) have not been stated. Various extensions of the differenced stopping rule have been proposed in

[13]Consider a decision-maker who receives a temporal sequence of independent, stochastic, signals about a binary state of the world, and must select an appropriate policy. The decision-maker must accumulate the imperfect state information until sufficient evidence about the current state yields an efficient decision. The Sequential Probability Ratio Test (Wald and Wolfowitz, 1948), which achieves a specified error rate with the smallest number of samples on average, is the normative solution to this problem. The DDM implements this solution exactly.

the literature (McMillen and Holmes, 2006; Krajbich and Rangel, 2011; Niwa and Ditterich, 2008; Ditterich and Churchland, 2012), however each of them requires determining which accumulator is greatest at every t then comparing it to other alternatives. From a computational standpoint, if the role of a BAM is to implement a "max" operator on subjective value, it would be (infinitely) recursive to embed a second "max" operation inside the stopping rule.

These issues have led researchers to explore alternative accumulation processes and stopping rules. In particular, there is recognition that modelling binary choice as resulting from a single accumulator to dual threshold arises from mathematical convenience, not from neurobiological plausibility (Kiani, Corthell, and Shadlen, 2014). The neurobiological evidence suggests multiple, competing accumulators $\mathbf{Z}(t)$ are the primitive objects underlying the neural processes of decision (Mazurek, Roitman, Ditterich, and Shadlen, 2003; Bogacz, Usher, Zhang, and McClelland, 2007; Beck, Ma, Kiani, Hanks, Churchland, Roitman, Shadlen, Latham, and Pouget, 2008; Churchland, Kiani, and Shadlen, 2008). In particular, the evidence is incompatible with a stopping rule which depends on the difference between accumulators, and instead points to a fixed threshold that is constant over trials (Roitman and Shadlen, 2002; Churchland, Kiani, and Shadlen, 2008; Niwa and Ditterich, 2008). The Race rule has this feature. Moreover, this form of stopping rule is easily generalizable to n choice alternatives.

A fixed stopping rule also has the feature that the "confidence" of a decision (the posterior belief that $v_1 > v_2$) can be inferred from the magnitude of $Z_1(t^*) - Z_2(t^*)$ at the time of decision (Drugowitsch and Pouget, 2012). Intriguingly, neural measurements of this quantity correlate with the decision to opt out of an uncertain choice decision in favour of a certain option (Kiani and Shadlen, 2009; Kiani, Corthell, and Shadlen, 2014). This observation is not compatible with a simple differenced stopping rule since

$Z_1(t^*) - Z_2(t^*) = \theta$ is constant by definition.

However for a simple Brownian motion accumulator, the Race rule carries with it the odd prediction that decision times do not depend on the magnitude difference between choice alternatives, an empirical regularity that has been observed in a wide range of experimental datasets (Luce, 1986). For this reason, applications of the Race rule which aim for neurobiological plausibility use the more general stochastic process given in (3) (Usher and McClelland, 2001; Tsetsos, Gao, McClelland, and Usher, 2012). The coupling of the accumulators is grounded in principles of neural computation (e.g. mutual inhibition Wang, 2002), and allows the distribution of decision time to be influenced by all of the choice alternatives.

It is important to emphasize that in all of the above implementations, the equation determining the choice is given by (7) and the difference between BAMs lies in the distribution of stopping times implied by the choice of accumulator and stopping rule. As we will see, the fact that all BAMs can be written with this choice criterion and an appropriate distribution of stopping times allows us to derive the equivalence between bounded accumulation models and the Fechner RUM.

3. DERIVATION OF STRONG RANDOM UTILITY MODEL

In an accumulation model, the probability that an accumulator reaches a particular threshold (thereby implementing a choice) depends on the vector $\mathbf{v}$. The open question is whether these choice probabilities can be represented by an additive form of the RUM. For the special case of the DDM, this task is straightforward because the choice probabilities can be derived in closed form: it is well-known that the DDM implies logistic choice probabilities.

PROPOSITION 1 *For $n = 2$, if $Z_i(t)$ is a Brownian motion (equation 5) and the stopping rule is differenced (equation 8), then the resulting choice* *probabilities can be represented by a Fechner RUM. Moreover, this RUM is of the Logit form.*

PROOF: $Z_1(t) - Z_2(t)$ is a one-dimensional Brownian motion to dual threshold (see Remark 1). Cox and Miller (1965) give the probability that this process hits the boundary for alternative i before j as

$$P_i(\mathbf{v}) = \left(1 + e^{\frac{-2(v_i - v_j)\theta}{\sigma^2}}\right)^{-1}. \tag{10}$$

Since the binary logistic choice probabilities are implied by a Fechner RUM with independent η_i distributed Gumbel (Luce and Suppes, 1965; McFadden, 1974), we have our result.[14] *Q.E.D.*

This result also highlights an identification issue that will be familiar to practitioners. If we assume $v_i - v_j$ is known, it is clear that only the term $\beta \equiv \frac{-2\theta}{\sigma^2}$ is identified by choice data. Therefore the variance of the stochastic process must be normalized before a Logit model can be applied.[15] However there is still something to be gained from the accumulation model. Incorporating a second observable – the decision time t^* – identifies an additional parameter of the DDM, resulting in a more efficient estimate of the relationship between $v_i - v_j$ and the choice probability than through choice data alone (Clithero and Rangel, 2014).

Unfortunately, for the general class of accumulation models and stopping rules given by equations (3) and (6), closed form expressions for the choice probabilities are not known. However we now demonstrate that this general model class can be reduced to a Fechner RUM in which the distribution

[14]However, for the binary the case, the reverse is not true. There is a class of possible distributions for $\boldsymbol{\eta}$ that is implied by the logistic choice probability (Yellott Jr., 1977). We will say more on this point in section 4.

[15]In many applications, the term $v_i - v_j$ is often assumed to be an unknown linear function of observables with parameter μ. Then $\beta \equiv \frac{-2\mu\theta}{\sigma^2}$.

of $\boldsymbol{\eta}$ can be characterized. Our result can be stated in the form of two propositions. First, for the case of binary choice and coupled accumulators:

PROPOSITION 2 *For $n = 2$, the accumulator given by (3) with $\boldsymbol{\Gamma}(t) = \psi\mathbf{1} + (\gamma - \psi)\mathbf{I}$, any stopping rule given by (6), and choice criterion given by (7), the resulting choice probabilities can be represented by a Fechner RUM.*

PROOF: In appendix 7.2. *Q.E.D.*

For the case of choice between n alternatives, our result can be shown for the case of uncoupled accumulators:[16]

PROPOSITION 3 *For $n \geq 2$, the accumulator given by (4), any stopping rule given by (6), and choice criterion given by (7), the resulting choice probabilities can be represented by a Fechner RUM.*

PROOF: In appendix 7.1. *Q.E.D.*

The derivation of a Fechner RUM from a BAM has two important implications for discrete choice modelling. First, the choice probabilities implied by a BAM can be captured within the empirical framework currently used by economists. Since the entire class of RUMs can be approximated by a GEV model (Dagsvik, 1995; McFadden and Train, 2000), this means that the choice probabilities resulting from a BAM can also be approximated by GEV with an appropriate covariance structure. Proposition 1 is an (exact) special case of this fact. For the general class of BAMs, it therefore remains for the choice modeller to specify the correct structure of the error term. This is where the second implication provides guidance. To provide some intuition for the derived distribution of $\boldsymbol{\eta}$, we present a proof of Proposition 3 for the special case of a Brownian-motion with drift (5).

[16] A derivation for coupled accumulators in discrete time can be found in a working paper (Webb, 2013).

Let us begin by noting that the choice criterion (7) is preserved under a linear scaling, $\lambda > 0$, at time $t^* > 0$.[17] Therefore,

$$i^* = \underset{i}{\operatorname{argmax}} \{Z_i(t^*)\} = \underset{i}{\operatorname{argmax}} \{\lambda Z_i(t^*)\}, \tag{11}$$

and equivalently for the choice probabilities,

$$P_{i^*} = \Pr\left[i^* = \underset{i}{\operatorname{argmax}} \{\lambda Z_i(t^*)\}\right].$$

Now, we solve the differential equation (5) at time t^*,

$$Z_i(t^*) = v_i t^* + \sigma B_i(t^*), \tag{12}$$

and $Z_i(t^*)$ is fully characterized by separable terms consisting of the exogenous value v_i and the realizations of the stochastic process $B_i(t)$ for $t \in [0, t^*]$.

Substituting (12) into (11) yields the choice criterion

$$i^* = \underset{i}{\operatorname{argmax}} \{\lambda v_i t^* + \lambda \sigma B_i(t^*)\}.$$

Now all that remains is to choose a suitable value for λ. If we define

$$\lambda \equiv \frac{1}{t^*} > 0, \tag{13}$$

the choice criterion becomes

$$i^* = \underset{i}{\operatorname{argmax}} \left\{v_i + \frac{\sigma}{t^*} B_i(t^*)\right\}.$$

This expression has the form of a Fechner RUM (1) in which η_i is comprised of the location of the Brownian motion $B_i(t^*)$, scaled by the stopping time t^*,

$$\eta_i = \frac{\sigma}{t^*} B_i(t^*), \tag{14}$$

[17] This is equivalent to observing that the scale of the random utility model is arbitrary, thus requires normalization.

with choice probabilities given by

$$P_i = \Pr\left[v_i - v_j > \tilde{\eta}_{ji}, \quad \forall j \neq i\right]$$
$$P_i = \Pr\left[v_i - v_j > \frac{\sigma}{t^*}\left(B_j(t^*) - B_i(t^*)\right), \quad \forall j \neq i\right]. \tag{15}$$

This derivation of η_i for the special case of Brownian motion clarifies the link between the stochastic process governing accumulation and the choice probabilities from the implied RUM. Propositions 2 and 3 state similar results for more general forms of accumulation within the class of BAM.

Of particular importance is the distribution of the t^* implied by the choice of accumulator and stopping rule. As the general class of BAMs is empirically constrained (whether via behavioural or neural data), this guides the choice of structure to impose on the behavioural model. Economic datasets which collect data on decision time will yield a more accurate characterization of this distribution, therefore also the distribution of $\tilde{\eta}_{ji}$. In some cases, the relationship between t^* and $\mathbf{v}$ arising from accumulation will allow the moments of $\tilde{\eta}_{ji}$ to be characterized in terms of $\mathbf{v}$. We know explore this relationship in greater detail.

4. THE DISTRIBUTION OF STOCHASTIC CHOICE

The distribution of stochastic choice that arises from a BAM depends on the form of accumulation (e.g. γ, ψ), the statistics of the stochastic process (e.g. σ), and the distribution of stopping time t^* given by the stopping rule. In particular, the presence of the the random variable t^*, in addition to the stochastic process $B_i(t)$, requires some consideration.

Some intuition can be gleaned from the simplified case of Brownian motion derived in (14). As previously noted, the random variable $\tilde{\eta}_{ji}$ can be expressed in terms of the stochastic processes $B(t)$, scaled by the stopping time t^*,

$$\tilde{\eta}_{ji} = \frac{\sigma\big(B_j(t^*) - B_i(t^*)\big)}{t^*}. \tag{16}$$

In fact, the re-expression of $\tilde{\eta}_{ji}$ as a ratio also holds for the more general processes (see sections 7.2 and 7.1). Though the ratio of random variables does not easily lend itself to characterization, the fact that the denominator increases in t^* will allow us to make statements about the distribution and/or moments of $\tilde{\eta}_{ji}$ for various cases of the general formulation.

In doing so, we will find it convenient to introduce the concept of a *first passage time*. For each accumulator $Z_i(t)$, the random variable t_i is defined as the first time $Z_i(t)$ exits the region defined by the stopping rule. For example, with a horse-race stopping rule, $t_i = \inf\{t : Z_i(t) \geq \theta_i\}$. In turn, the decision time is denoted by $t^* \equiv \min\{t_1, \ldots, t_n\}$. We denote the CDF of the first passage time distribution by $G_i(t; \mathbf{v}, \mathbf{\Gamma}(t), \boldsymbol{\sigma}(t))$ where the dependence on the (functional) parameters of the accumulation process is explicit. When this detail is not required, we denote the CDF $G_i(t)$ with density $g_i(t)$.

In the literature on stochastic processes, first passage times are, arguably, the primary objects of mathematical investigation. In the current application, they will prove useful for three reasons.

First, if we restrict the accumulator to Brownian motion, first passage times allow a direct characterization of the choice probabilities in closed form. In Proposition 1, we observed that a Brownian motion accumulator combined with the Differenced stopping rule (i.e. the DDM) yields logistic choice probabilities. This result is derived, in part, from the first passage time distributions (Cox and Miller, 1965). In section 4.1, we verify that the formulation of $\tilde{\eta}_{ji}$ resulting from BAM recovers the Logit model from a wider class of distributions for $\boldsymbol{\eta}$ than previously stated in the literature.

Second, the first passage times provide a means to characterize t^* by virtue of the stopping rule. A useful example is the Race stopping rule, where the characterization of t^* is relatively straightforward. In section 4.2, we pursue this example to demonstrate how the distribution of t^*, and its dependence on $\mathbf{v}$, can be characterized for the class of accumulators given by

(4). By means of Proposition 3, we can then draw conclusions on how the distribution of the implied Fechner RUM depends on $\mathbf{v}$.

Third, for the general class of accumulators in (3), where no closed form expressions for the distributions of $Z(t)$, t^*, or even t_i exist, numerical approximations of the first passage time distributions are still possible. Through sampling from these distributions, stopping rules given by equation (6) can be implemented via simulation, yielding simulated choice probabilities. We describe these methods in section 4.3 before moving to an example application in section 5.

4.1. *Differenced stopping rule*

We begin by providing some intuition for how Proposition 3 relates to Proposition 1. Under a Differenced stopping rule (8), the magnitude of the difference between the accumulators at t^* must be fixed at θ. Therefore

$$Z_i(t^*) - Z_j(t^*) = R,$$

and

$$B_i(t^*) - B_j(t^*) = \frac{R - t^*(v_i - v_j)}{\sigma}, \tag{17}$$

where R is a random variable which takes on θ if the upper threshold is crossed (alternative i is chosen), and $-\theta$ otherwise. For intuition, consider the case in which $v_i - v_j = 0$ and $\sigma = 1$; then $B_j(t^*) - B_i(t^*)$ must either be θ or $-\theta$ for the accumulation to terminate (Figure 1).

Substituting (17) into (16) (noting the subscript i and j) then yields

$$\tilde{\eta}_{ji} = \frac{-R}{t^*} + (v_i - v_j),$$

with choice probabilities given by

$$\begin{aligned} P_i(\mathbf{v}) &= \Pr\left[v_i - v_j > \frac{-R}{t^*} + (v_i - v_j), \quad \forall j \neq i\right] \\ &= \Pr\left[\frac{R}{t^*} > 0, \quad \forall j \neq i\right]. \end{aligned}$$

Noting that $t^* > 0$, since the probability that $R = \theta > 0$ is given by the logistic function in equation (10), we have therefore recovered the Logit model, but from a wider class of distributions for $\boldsymbol{\eta}$ then previously noted in the literature (i.e. the independent Gumbel; Luce and Suppes, 1965; McFadden, 1974). There are two issues to note.

First, the necessity result typically stated in the literature – linking logistic choice probabilities to the Gumbel distribution for η_i – requires independence (McFadden, 1974), and only holds for cases in which there are more than two choice alternatives (Yellott Jr., 1977). The first condition does not hold in the case of the DDM since η_i and η_j are not independent through their relation to t^*. Moreover, the existing closed form results for the DDM can only be applied in the binary case. Therefore it is not required that the derived distribution for $\boldsymbol{\eta}$ be Gumbel.

Second, the distribution of $\boldsymbol{\eta}$ (or $\tilde{\eta}_{ji}$) derived from the DDM is particularly unusual. While $t^* > 0$ scales the random variable R, it has no bearing on whether the ratio $\frac{R}{t^*}$ is greater or less than 0. Therefore it is irrelevant in terms of determining the distribution. It turns out this property is unique to the Differenced stopping rule, as we will see in the following section.

Beyond the Brownian motion process, closed forms for the choice probabilities under the Differenced stopping rule are not available. For the O-U process, Busemeyer and Townsend (1992) give an expression in terms of a ratio of definite integrals. Numerical simulations suggest these probabilities indeed are of the logistic form. For the general accumulation process (3), no analytic statements for the choice probabilities are available.

However for binary choice, the choice probabilities can be approximated via numerical methods. If $\bar{t}$ is defined as the first stopping time at the upper boundary for the 1-D process $Z_1(t) - Z_2(t)$, conditional on not having crossed the lower, then the first passage time density $g_{\bar{t}}(t)$ can be approximated by methods given in Smith (2000).[18] The choice probability for alternative 1 is then given by integrating the first passage time density via numerical methods,

$$P_1(\mathbf{v}) = \int_0^\infty g_{\bar{t}}(t)\, dt.$$

Conveniently, these methods also apply to common extensions of the differenced stopping rule, in particular a threshold $\theta(t)$ which depends on time (Drugowitsch, Moreno-Bote, Churchland, Shadlen, and Pouget, 2012). Numerical simulations suggest that these probabilities deviate from the Logit. Finally, if a dataset also contains information on decision times, likelihood methods can be used to incorporate this added information to yield a more efficient estimate via,

$$P_1(\mathbf{v}, t^*) = \int_0^{t^*} g_{\bar{t}}(t)\, dt.$$

Clithero and Rangel (2014) given an empirical demonstration of this result in the case of the DDM.

4.2. *Race stopping rule*

A Race stopping rule is an example of a stopping rule in which the levels, not just the differences, of each accumulator are relevant to the decision. Importantly, it allows an easy characterization of the stopping time $t^*(\mathbf{v}) = \min \mathbf{t}(\mathbf{v}) = \min[t_1(v_1), \ldots, t_n(v_n)]$. Combined with the expression for $\tilde{\eta}_{ji}$ derived in Section 3, this will allow us to state results for the moments of $\tilde{\eta}_{ji}$, and how they depend on $\mathbf{v}$, for the accumulation process in (4).

PROPOSITION 4 *For all uncoupled processes $Z_i(t)$, $i = 1 \ldots n$, with $\sigma(t) = \sigma$, and $E[t_i]$ which decreases in v_i, the resulting random variable $\tilde{\eta}_{ji}$ (via*

[18] Computational methods for implementing these approximations are available under a BSD license at https://github.com/jdrugo/dm

Proposition 3) has mean 0 and variance which increases in any element of $\mathbf{v}$.

PROOF: Proof is given in appendix 7.3. *Q.E.D.*

The intuition for Proposition 4 is straight-forward. Any increase in the value v_i, for any process $Z_i(t)$, will lower the expected time that the process will terminate, therefore lower the expected time of the decision generally. Since the denominator of $\tilde{\eta}_{ji}$ is increasing in t^* (e.g. equation **??**), then any increase in the value v_i for any alternative will reduce the denominator, increasing the variance of $\tilde{\eta}_{ji}$.

In the case of Brownian motion and a Race stopping rule, the conditions for Proposition 4 are easy to verify. It is well known that the stopping time of a Brownian motion (to a single threshold) follows the Inverse Gaussian distribution with mean $\frac{\theta}{v_i}$ and variance $\frac{\theta^2}{\sigma^2}$ (e.g. Cox and Miller, 1965, p. 221). Conditions for the O-U process which satisfy Proposition 4 are given in Busemeyer and Townsend (1992) and Smith (2000). In fact, intuition suggests that Proposition 4 is generalizable to the accumulation in equation 3, provided the variance of its solution is of order less than t^2. Simulation for binary choice under the conditions used in Proposition 2 suggests this is indeed the case (Usher and McClelland, 2001). However, since the Differenced stopping rule does not satisfy the requirement that $E[t_i]$ decreases in v_i, Proposition 4 does not apply in this case even for general accumulation processes.

The result of Proposition 4 comes with an interesting implication for choice sets in which the magnitude of the subjective values are scaled in such a way that their relative differences are preserved.

COROLLARY 1 *If the conditions of Proposition 4 are satisfied, then for all* $i \neq j$ *such that* $v_i - v_j > 0$, $P_i(\mathbf{v}) > P_i(\mathbf{v}+\boldsymbol{\alpha}) > \frac{1}{n}$ *for all* $\boldsymbol{\alpha} = [\alpha, \ldots, \alpha] > \mathbf{0}$.

PROOF: The choice probabilities are given by equation (2), restated here as an integral over the set of realizations of $\boldsymbol{\eta}$ for which i is chosen.

$$P_i(\mathbf{v}) = \int_{\boldsymbol{\eta} \in D_{\mathbf{v}}} f_{\mathbf{v}}(\boldsymbol{\eta})\, d\boldsymbol{\eta},$$

where

$$D_{\mathbf{v}} = \{\boldsymbol{\eta} : v_i - v_j > \tilde{\eta}_{ji}, \quad \forall j \neq i\},$$

and $f()$ is the joint density of $\boldsymbol{\eta}$ implied by the BAM with valuation $\mathbf{v}$. Since $\mathbf{v} + \boldsymbol{\alpha}$ preserves the relative differences between alternatives, the elements of the set B are identical under $\mathbf{v}$ and $\mathbf{v} + \boldsymbol{\alpha}$:

$$D_{\mathbf{v}+\boldsymbol{\alpha}} = \{\boldsymbol{\eta} : v_i + \alpha - v_j - \alpha > \tilde{\eta}_{ji}, \quad \forall j \neq i\} = D_{\mathbf{v}}.$$

From Proposition 4, $f_{\mathbf{v}+\boldsymbol{\alpha}}(\boldsymbol{\eta})$ transfers probability density from the centre of the (mean zero) distribution towards the tails. In particular, for all $i \neq j$ such that $v_i - v_j > 0$, density is transferred towards realizations of $\boldsymbol{\eta} \notin D$. Therefore $P_i(\mathbf{v}) > P_i(\mathbf{v} + \boldsymbol{\alpha})$. Moreover, as $\alpha \to \infty$, $P_i(\mathbf{v} + \boldsymbol{\alpha}) = \frac{1}{n}$. *Q.E.D.*

Corollary 1 states that the choice probability of the highest-valued alternative (and logically, the other alternatives as well) depends on both the differences in subjective valuations *and* the magnitudes. A Fechner RUM with constant variance, such as the Logit, does not share this property.

It is important to note that the scaling of the variance with $\mathbf{v}$ does not imply that preferences (in a stochastic sense) are re-ordered. If $P_i(\mathbf{v}) > P_j(\mathbf{v})$, then $P_i(\mathbf{v} + \boldsymbol{\alpha}) > P_j(\mathbf{v} + \boldsymbol{\alpha})$. Moreover, the increase in variance of $\tilde{\eta}_{ji}$ from increasing any v_i does *not* reduce the probability that i is chosen. We can state this result explicitly for the case of a Brownian motion accumulator.

PROPOSITION 5 *For a Brownian motion accumulator (5) and a Race stopping rule (9), then* $P_i(\mathbf{v}\boldsymbol{\alpha}^\top) > P_i(\mathbf{v})$ *for some* $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]$ *where* $\alpha_i > 1$ *and* $\alpha_j = 1, \forall j \neq i$.

A proof is given in appendix 7.4. In terms of the BAM, the argument is intuitive. Any increase in the value of an alternative will increase the probability that its accumulator will hit the threshold first, regardless of the stochasticity added by the decrease in the expected stopping time.

The dependence between choice stochasticity and the magnitude of valuations carries with it an important implication. In the absence of some mechanism for re-scaling valuations, the Race stopping rule implies more stochasticity in choice behaviour for more valuable choice sets than predicted by a model with constant variance. From the perspective of optimal choice, this suggests that the brain might have some mechanism to re-scale, or normalize, valuations relative to a given threshold, reducing the number of errors. In a recent article, we explore such a mechanism that results from known neurobiological constraints (Webb, Glimcher, and Louie, 2014).

4.3. *Numerical Methods for Approximating Choice Probabilities*

Where closed form expressions for the distributions of $Z(t)$, t^*, or t_i do not exist, as in the general accumulation (3), a numerical method for approximating the first passage time density is still possible. To demonstrate, we pursue an example using the Race stopping rule since it allows easy characterization of the stopping time t^*, and it will be useful in our empirical example in section 5. Note that this method generalizes to other stopping rules, including the Differenced rule as well as rules not pursued in this article. All that is required is a functional relation deriving the choice i^* from the first passage times $\{t_1, \dots, t_n\}$.

In the case of the Race stopping rule, this function is given by $i^* = \arg\min\{t_1, \dots, t_n\}$. One immediate implication is that the random variables t_i can be directly interpreted as a random utility,

$$i^* = \underset{i}{\text{argmax}}\left\{u_i\right\} = \underset{i}{\text{argmax}}\left\{Z_i(t^*)\right\} = \underset{i}{\text{argmin}}\left\{t_i\right\},$$

since the random vector $[t_1, \ldots, t_n]$ represents the choice probabilities (Marley and Colonius, 1992). Therefore,

$$(18) \qquad P_i = \Pr\left[t_i < t_j, \quad \forall i \neq j\right].$$

Since the density $g_i(t; \mathbf{v})$ can be approximated numerically (Smith, 2000), repeated sampling from this density yields a simulated approximation of P_i and the methods of Maximum Simulated Likelihood can be applied (Train, 2009). We now pursue an example of this method, and the results from the earlier sections, in the case of choice under uncertainty.

5. EXAMPLE: MEASURING RISK AVERSION

The link between a BAM and a Fechner RUM carries with it a range of implications for economic modelling. As noted in the previous section, these include hypotheses for how the distribution of stochastic choice depends on $\mathbf{v}$, and methods for incorporating additional data on decision time into an econometric specification. To give a simple example, we now consider the estimation of a structural parameter for risk aversion from a well-known experiment on choice over uncertainty (Holt and Laury, 2002).

In the Holt and Laury experiment, each subject was presented with a list of lottery pairs (Table I), ordered by the difference in their expected value. This method is commonly referred to as a 'Multiple Price List' experiment. As each subject picks a lottery from each of the pairs, their willingness to choose a lottery with negative expected value reflects the degree of risk aversion in their preferences. From this choice data, it is therefore possible to estimate a utility curvature parameter α from a CRRA utility function $u(x) = x^{\alpha}$, using standard likelihood methods (Camerer and Ho, 1994; Hey and Orme, 1994; Holt and Laury, 2002; Harrison and Rutstrom, 2008), provided one places an assumption on the distribution of stochastic choice. Our goal is to explore the implications that misspecification of the stochastic choice

TABLE I

LOTTERY CHOICES IN THE HOLT-LAURY RISK AVERSION EXPERIMENT (REPRODUCED FROM HARRISON AND RUTSTROM, 2008)

Lottery 1 (Safe)				Lottery 2 (Risky)				EV_1	EV_2	EV_1-EV_2
$p_{\$2}$		$p_{\$1.60}$		$p_{\$3.85}$		$p_{\$0.10}$				
0.1	$2	0.9	$1.60	0.1	$3.85	0.9	$0.10	$1.64	$0.48	$1.17
0.2	$2	0.8	$1.60	0.2	$3.85	0.8	$0.10	$1.68	$0.85	$0.83
0.3	$2	0.7	$1.60	0.3	$3.85	0.7	$0.10	$1.72	$1.23	$0.49
0.4	$2	0.6	$1.60	0.4	$3.85	0.6	$0.10	$1.76	$1.60	$0.16
0.5	$2	0.5	$1.60	0.5	$3.85	0.5	$0.10	$1.80	$1.98	−$0.17
0.6	$2	0.4	$1.60	0.6	$3.85	0.4	$0.10	$1.84	$2.35	−$0.51
0.7	$2	0.3	$1.60	0.7	$3.85	0.3	$0.10	$1.88	$2.73	−$0.84
0.8	$2	0.2	$1.60	0.8	$3.85	0.2	$0.10	$1.92	$3.10	−$1.18
0.9	$2	0.1	$1.60	0.9	$3.85	0.1	$0.10	$1.96	$3.48	−$1.52
1	$2	0	$1.60	1	$3.85	0	$0.10	$2.00	$3.85	−$1.85

distribution might yield for the estimate of risk aversion in this experiment, particularly if choice results from a dynamic decision process.

We do so via Monte Carlo methods. Lottery choices were simulated for twenty risk-neutral subjects ($\alpha = 1$), with choice stochasticity introduced using the BAM. The accumulator was parameterized as a Brownian motion (equation 5), with the valuation of lottery i given by $v_i = EU_i$, and the decision terminated via the Race stopping rule with $\theta = 10$. The variance of the accumulation, σ, was normalized to 1.

Under a Race stopping rule, the relationship between the magnitude of valuations and the variance of $\tilde{\eta}_{ji}$ is subtle. As noted in Proposition 4, the variance of $\tilde{\eta}$ depends on the magnitude of v_i, specifically on the expected utility of the safe and risky lottery, but not so much as to decrease choice probabilities (Proposition 5). This implies that the variance of $\tilde{\eta}_{ji}$ *depends on the lottery pair* in a Multiple Price List experiment. Econometrically, this amounts to a Fechner RUM with heteroskedasticity of a form that, if ignored in estimation, will lead to misspecification of the choice probabilities and biased estimates of model parameters.

Indeed, Table II reports the average estimate, $\hat{\alpha}$, from 1000 simulated datasets, estimated under the common assumption of logistic choice probabilities and a constant standard deviation s. In terms of equation 11, we denote this specification H_0,

$$H_0 : \quad \frac{\sigma^2}{2\theta} = s,$$

therefore $P_i = \left(1 + e^{\frac{EU_1 - EU_2}{s}}\right)^{-1}$.

Under the specification of constant variance, the average estimate $\hat{\alpha}$ is 0.96, with a Type I error rate of 10.1% given a 5% significance-level test. The bias in the estimate of $\hat{\alpha}$ arises because the variance of $\tilde{\eta}_{ji}$ depends, in part, on the magnitude of the expected utilities of each of the lotteries. In the Holt and Laury experiment, the lotteries with the largest expected utilities are found in pairs where $EU_1 - EU_2 < 0$, increasing the choice stochasticity for these lotteries. As a result, the choice probabilities, as a function of $EU_1 - EU_2$, are not symmetric around zero (as implied by a logistic function) and the estimator attempts to compensate by biasing the estimate of α downwards (Figure 2). Since α enters the utility function non-linearly, the degree of bias varies depending on the lottery, but it is consequential. For the 50/50 version of Lottery 2, it amounts to a risk premium of $\sim 5\%$ of the expected value. For a 50/50 lottery between \$0 and \$100, the risk premium would amount to $\sim 17\%$.

Fortunately, a range of methods are available to achieve less biased (or even unbiased) results, depending on the specification of choice stochasticity and the availability of decision time data. Since the variance of $\tilde{\eta}_{ji}$ depends on $\mathbf{v}$, one possibility is to include the difference in expected utility in the

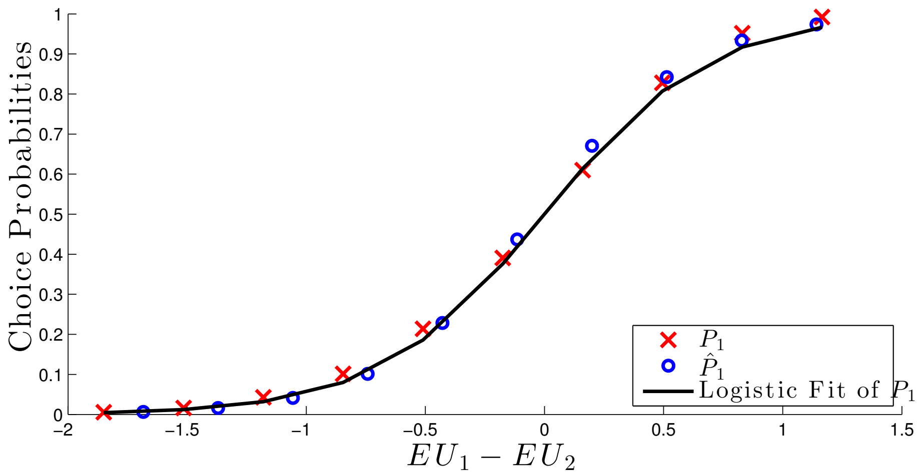

FIGURE 2.— The true probability of choosing lottery 1 (denoted $\times$) resulting from a BAM with Race stopping rule, as a function of the difference in EU between the lotteries 1 and 2. For comparison, the probabilities implied by the biased estimate of $\hat{\alpha}$ ($\circ$) are depicted, as well as a logistic fit of P_1 by a model with constant variance.

TABLE II

Average estimates of the CRRA coefficient from 1000 simulations of a risk-neutral chooser ($\alpha = 1$), for various specifications of the stochastic choice distribution.

	$\hat{\alpha}$	Type I Error Rate	$\hat{s}^2$	$\hat{k}$	$\hat{h}_1$	$\hat{h}_2$	$\hat{g}$
H_0	0.960	0.101	0.101				
H_1	0.973	0.076	0.032	-0.368			
H_2	0.981	0.071	0.039	-0.216	-6.784	-0.066	
H_3	0.978	0.085	0.184				-.014
H_4	0.997	0.053	0.0414				

specification,

$$H_1 : \quad \frac{\sigma^2}{2\theta} = s\sqrt{e^{k(EU_1 - EU_2)}},$$

as well as the magnitudes,

$$H_2 : \quad \frac{\sigma^2}{2\theta} = s\sqrt{e^{k(EU_1 - EU_2)} + e^{h_1 EU_1} + e^{h_2 EU_2}}.$$

As can be observed in Table II, including functions of $\mathbf{v}$ reduce the bias of $\hat{\alpha}$ by nearly half and reduce Type I errors to 7.1%. However, these linear specifications do not fully capture the dependence of the choice probabilities on $\mathbf{v}$, since the bias is not completely eliminated.

Another possibility is to use data on decision time directly. Since $\mathbf{v}$ directly impacts the distribution of stopping times, the effect on the variance of the random utility model can be equivalently stated in terms of time: the variance of $\tilde{\eta}_{ji}$ increases as expected decision times becomes shorter. This leads to the following specification,

$$H_3 : \quad \frac{\sigma^2}{2\theta} = \sqrt{s^2 + gt^*}.$$

Including stopping time data has a similar effect to including a linear function of $\mathbf{v}$, reducing the bias in $\hat{\alpha}$ by half and the Type I error rate to 8.5%.[19]

[19]To ensure this linear specification of variance is greater than zero, a cutoff rule was

Finally, if the choice probabilities are correctly specified, the methods in section 4.3 can be used to arrive at estimates for α and θ, even if data on decision times are not available. As noted in section 4.2, the Race stopping rule implies that the stopping times are distributed Inverse Gaussian with mean $\frac{\theta}{EU_i}$ and variance θ^2. Drawing a large number of samples from these distributions (for each lottery) yields an approximation of P_i via equation (18) for a given α and θ. We denote this specification H_4, with θ reported in Table II as $\frac{1}{2s^2}$ to conform with the other specifications. As expected, a Maximum Simulated Likelihood estimation of α appears unbiased, with a Type I error rate of 5.3%.

6. CONCLUSION

Over the past fifteen years, significant progress has been made in understanding the neural processes underlying choice, with particular focus on the dynamics of a decision. From this line of theoretical and empirical research, a class of models, termed Bounded Accumulation Models, have been developed to link both behavioural and neural data. Here, we demonstrate that a "Strong" Random Utility Model – the benchmark framework for discrete choice in economic application – can be derived from a general class of BAMs, thereby providing a neurobiological foundation for random utility maximization.

The relationship between these two approaches might not be surprising given the origins of the RUM in psychophysical research (McFadden, 2001). However the implications of this relationship has important consequences for testing economic theory and predicting choice behaviour. We demonstrate that the particular parameterization of a BAM – particularly its dynamics and stopping rule – influence the resulting distribution of stochastic choice:

used with the likelihood heavily penalized. Notably, the non-linear specification $\frac{\sigma^2}{2\theta} = se^{gt^*}$ did not reduce bias.

our derivation of a RUM implies that distribution of stochastic choice depends on the distribution of stopping times implied by a BAM.

In the special case of a Brownian motion to dual thresholds (better known as a Drift Diffusion model for binary choice), the resulting choice distribution is logistic, though we derive it from a more general class of error distributions than traditionally reported in the literature. More neurobiologically-plausible models imply that the distribution of choice will depend on observables. To demonstrate this, we pursue an example of the Race stopping rule (for multiple alternatives) in which the distribution of stopping times is known. Under this rule, the variance of the choice distribution will increase in the magnitude of the valuation of each of the alternatives, or equivalently, decrease in the observed decision times. As a result, estimates of structural choice parameters can be biased if the stochastic choice distribution is misspecified using common techniques. To partially correct for this bias, we introduce specifications of the variance which depend on both the observables of choice alternatives and decision time data (if observed). To fully correct for the bias, we present a method for approximating the choice probabilities through simulating the stopping time distributions, though this does require the correct specification of the BAM. Conveniently, these methods apply for the general formulation of a BAM (not just the “Race” rule), since numerical approximations of the stopping time distributions exist.

This article also points to a methodological goal. Knowledge of the choice process in the brain will constrain behavioural models (Bernheim, 2009; Webb, 2011). Clearly, much more work is left to be done before a sharper specification of the stochastic choice distribution is possible, including elements that lie outside the brain but within a standard applied model. However the mathematical relationship between these two levels of analysis implies that advances in neuroscientific modelling will reduce the degrees of freedom that applied economics researchers must consider, and vice-versa.

In essence, modelling at the level of behaviour and at the level of dynamic neural processes is a symbiotic exercise.

7. APPENDIX

7.1. *Derivation of Fechner RUM for Uncoupled Processes*

Let us begin by noting that the choice criterion (equation 7) is preserved under a scaling, $\Lambda(t^*) > 0$, for $t^* > 0$. Therefore,

$$i^* = \underset{i}{\operatorname{argmax}} \left\{ Z_i(t^*) \right\} = \underset{i}{\operatorname{argmax}} \left\{ \Lambda(t^*) Z_i(t^*) \right\},$$

The Îto calculus can be used to derive a solution to (4) (Smith, 2000). Solving this equation at t^* yields

$$\begin{aligned} Z_i(t^*) &= \int_0^{t^*} \frac{U(t^*)}{U(\tau)} \big[\left[v_i + \nu(\tau) \right] \, d\tau + \sigma(\tau) \, dB_i(\tau) \big], \\ &= v_i \int_0^{t^*} \frac{U(t^*)}{U(\tau)} \, d\tau + \int_0^{t^*} \frac{U(t^*)}{U(\tau)} \big[\nu(\tau) \, d\tau + \sigma(\tau) \, dB_i(\tau) \big], \end{aligned} \tag{19}$$

where

$$U(t) = e^{\int_0^t \gamma(\tau) \, d\tau} > 0, \quad \forall t > 0.$$

Following the intuition from our previous example, we divide through equation (19) by $\Lambda(t^*) \equiv \left[U(t^*) \int_0^{t*} U^{-1}(\tau) \, d\tau \right]^{-1} > 0$, yielding

$$\Lambda(t^*) Z_i(t^*) = v_i + \Lambda(t^*) \int_0^{t^*} \frac{U(t^*)}{U(\tau)} \big[\nu(\tau) \, d\tau + \sigma(\tau) \, dB_i(\tau) \big].$$

Substituting this term into the choice criterion,

$$\begin{aligned} i^* &= \underset{i}{\operatorname{argmax}} \left\{ \Lambda(t^*) Z_i(t^*) \right\}, \\ &= \underset{i}{\operatorname{argmax}} \left\{ v_i + \eta_i \right\}, \end{aligned}$$

yields our result with

$$\eta_i \equiv \frac{\int_0^{t^*} U^{-1}(\tau) \big[\nu(\tau) \, d\tau + \sigma(\tau) \, dB_i(\tau) \big]}{\int_0^{t^*} U^{-1}(\tau) \, d\tau}. \tag{20}$$

Note that the random variable relevant for choice behaviour, $\tilde{\eta}_{ji}$, is thus given by

$$\tilde{\eta}_{ji} = \frac{\int_0^{t^*} U^{-1}(\tau)\sigma(\tau)\, d(B_j(\tau) - B_i(\tau))}{\int_0^{t^*} U^{-1}(\tau)\, d\tau}, \tag{21}$$

since the the urgency term $\nu(\tau)$ is not specific to alternatives i and j, therefore cancels.

7.2. *Derivation of Fechner RUM for Coupled Processes*

We begin with the solution to the differential equation (3), for $\mathbf{\Gamma}(t) = \mathbf{\Gamma} = \left(\begin{smallmatrix} \gamma & \psi \\ \psi & \gamma \end{smallmatrix}\right)$, as given in Smith (2000) and **?**. Let the $n \times n$ matrix function $\mathbf{\Delta}(t)$ be the fundamental solution to the first-order linear differential system

$$\mathbf{\Delta}'(t) = \mathbf{\Gamma}\mathbf{\Delta}(t), \quad \mathbf{\Delta}(0) = \mathbf{I}.$$

Then the solution to equation (3) at t^* is

$$\begin{aligned}
\mathbf{Z}(t^*) &= \mathbf{\Delta}(t^*) \int_0^{t^*} \mathbf{\Delta}^{-1}(\tau) \left[\mathbf{v} + \nu(\tau)\right] \, d\tau + \mathbf{\Delta}(t^*) \int_0^{t^*} \mathbf{\Delta}^{-1}(\tau) \boldsymbol{\sigma}(\tau)\, d\mathbf{B}(\tau) \\
&= \int_0^{t^*} \mathbf{\Delta}(t^*)\mathbf{\Delta}^{-1}(\tau)\mathbf{v}\, d\tau + \int_0^{t^*} \mathbf{\Delta}(t^*)\mathbf{\Delta}^{-1}(\tau) \left[\nu(\tau)\, d\tau + \boldsymbol{\sigma}(\tau)\, d\mathbf{B}(\tau)\right] \\
&= \int_0^{t^*} \mathbf{\Delta}(t^*)\mathbf{\Delta}^{-1}(\tau)\mathbf{v}\, d\tau + \mathbf{C}(t^*),
\end{aligned}$$

where for exposition we define the portion of the integral that does not depend on $\mathbf{v}$ as $\mathbf{C}(t^*)$.

To proceed, observe that the solution $\mathbf{\Delta}(t)$ takes the form

$$\mathbf{\Delta}(t) = \frac{1}{2} e^{\gamma t} \begin{pmatrix} \Psi_1(t) & \Psi_2(t) \\ \Psi_2(t) & \Psi_1(t) \end{pmatrix}, \tag{22}$$

where $\Psi_1(t) = e^{\psi t} + e^{-\psi t} > 0$, and $\Psi_1(t) = e^{\psi} - e^{-\psi t} > 0$, with its inverse given by

$$\mathbf{\Delta}^{-1}(t) = 2e^{-\gamma t} \begin{pmatrix} \Psi_1(t) & -\Psi_2(t) \\ -\Psi_2(t) & \Psi_1(t) \end{pmatrix}.$$

Therefore

$$\mathbf{\Delta}(t^*)\mathbf{\Delta}^{-1}(\tau) = e^{\gamma(t^*-\tau)} \begin{pmatrix} \bar{\Psi}_1(t,\tau) & \bar{\Psi}_2(t,\tau) \\ \bar{\Psi}_2(t,\tau) & \bar{\Psi}_1(t,\tau) \end{pmatrix},$$

where $\bar{\Psi}_1(t,\tau) = 2e^{\psi(t-\tau)} + 2e^{-\psi(t-\tau)} > 0$ and $\bar{\Psi}_1(t) = 2e^{\psi(t-\tau)} - 2e^{-\psi(t-\tau)} > 0$, for all t and τ.

Substituting this term in to the solution to the differential equation yields

$$\begin{aligned} \mathbf{Z}(t^*) &= \int_0^{t^*} e^{\gamma(t^*-\tau)} \left[\bar{\Psi}_1(t^*,\tau)\mathbf{v} + \bar{\Psi}_2(t^*,\tau) \begin{pmatrix} v_2 \\ v_1 \end{pmatrix} \right] d\tau + \mathbf{C}(t^*) \\ &= \mathbf{v} \int_0^{t^*} e^{\gamma(t^*-\tau)} \bar{\Psi}_1(t^*,\tau)\, d\tau + \int_0^{t^*} e^{\gamma(t^*-\tau)} \bar{\Psi}_2(t^*,\tau) \begin{pmatrix} v_2 \\ v_1 \end{pmatrix} d\tau + \mathbf{C}(t^*), \end{aligned}$$

where we now explicitly separate the portion of the accumulation of Z_i which depends on v_i from the portion which depends on the other alternative(s). From this point, we proceed as above. Define $\Lambda(t^*) \equiv \left[\int_0^{t^*} e^{\gamma(t^*-\tau)} \bar{\Psi}_1(t^*,\tau)\, d\tau\right]^{-1} > 0$, and multiply through, yielding

$$\Lambda(t^*)\mathbf{Z}(t^*) = \mathbf{v} + \Lambda(t^*) \int_0^{t^*} e^{\gamma(t^*-\tau)} \bar{\Psi}_2(t^*,\tau) \begin{pmatrix} v_2 \\ v_1 \end{pmatrix} d\tau + \Lambda(t^*)\mathbf{C}(t^*).$$

Finally, substituting into the choice criterion yields

$$\begin{aligned} i^* &= \operatorname{argmax} \{\Lambda(t^*)\mathbf{Z}(t^*)\} \\ &= \operatorname{argmax} \{\mathbf{v} + \boldsymbol{\eta}\}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\eta} &\equiv \Lambda(t^*) \int_0^{t^*} e^{\gamma(t^*-\tau)} \bar{\Psi}_2(t^*,\tau) \begin{pmatrix} v_2 \\ v_1 \end{pmatrix} d\tau + \Lambda(t^*)\mathbf{C}(t^*) \\ &= \frac{\int_0^{t^*} e^{-\gamma\tau} \bar{\Psi}_2(t^*,\tau)\, d\tau}{\int_0^{t^*} e^{-\gamma\tau} \bar{\Psi}_1(t^*,\tau)\, d\tau} \begin{pmatrix} v_2 \\ v_1 \end{pmatrix} + \Lambda(t^*)\mathbf{C}(t^*). \end{aligned}$$

7.3. *Results for the moments of $\tilde{\eta}_{ji}$ under a Race stopping rule*

Before a proof of Proposition 4 we establish the following two Lemmas.

LEMMA 1 *Given a function $f(x) : \mathbb{R}_+ \to \mathbb{R}_+$, the ratio $\frac{\int_0^a f(x)\,dx}{\sqrt{\int_0^a f^2(x)\,dx}}$ is increasing in a.*

PROOF: Consider the limit definition of the integral,

$$\int_0^a f(x)\,dx = \lim_{h\to\infty} \sum_{n=1}^{h} f(c_i)\Delta x,$$

where $c_n = \frac{a}{h}n$ and $\Delta x = \frac{a}{h}$.
Define a vector $\mathbf{f}_h \in \mathbb{R}_+^h$ with each element given by $f(c_i)$. Then the integral can be re-written in terms of the vector,

$$\lim_{h\to\infty} \sum_{i=1}^{h} f(c_n)\Delta x = \lim_{h\to\infty} \|\mathbf{f}_h\|_1 \Delta x,$$

where $\|\cdot\|_p$ is a norm of degree p.
Now stating the ratio in this form,

$$\begin{aligned}\frac{\int_0^a f(x)\,dx}{\sqrt{\int_0^a f^2(x)\,dx}} &= \frac{\lim_{h\to\infty}\sum_{i=1}^{h} f(c_i)\Delta x}{\lim_{h\to\infty}\sqrt{\sum_{i=1}^{h} f^2(c_i)\Delta x}} \\ &= \lim_{h\to\infty} \frac{\|\mathbf{f}_h\|_1 \Delta x}{\|\mathbf{f}_h\|_2 \sqrt{\Delta x}} \\ &= \lim_{h\to\infty} \frac{\|\mathbf{f}_h\|_1}{\|\mathbf{f}_h\|_2}\sqrt{\Delta x}.\end{aligned} \tag{23}$$

For a given a, $\|\mathbf{f}_h\|_1 > \|\mathbf{f}_h\|_2$, $\forall h$. Moreover, for $a' = a + \frac{a}{h}$, and $\mathbf{f}_{h+1} \equiv [\mathbf{f}_h, f(a')] \in \mathbb{R}_+^{h+1}$,

$$\frac{\|\mathbf{f}_{h+1}\|_1}{\|\mathbf{f}_{h+1}\|_2} > \frac{\|\mathbf{f}_h\|_1}{\|\mathbf{f}_h\|_2}.$$

Therefore the ratio given in (23) is increasing in a. *Q.E.D.*

LEMMA 2 *Consider the random variables X and Y, where $E[X] = 0$ and Y has support $(0, \infty]$. For an increasing function $g()$,*

$$Var\left(\frac{X}{g(Y)}\right) \approx \frac{Var(X)}{g^2(E[Y])}$$

via a Taylor expansion of $\frac{X}{g(Y)}$ around $(E[X], E[Y])$.

PROOF: We follow (Ord and Stewart, 1994, p. 351) but focus on the case $f(X,Y) = \frac{X}{g(Y)}$. The expectation of the first-order Taylor expansion of $f(X,Y)$ around $\theta = (E[X], E[Y])$ is given by

$$\begin{aligned} E[f(X,Y)] &\approx E[f(\theta)] + E[f'_X(\theta)(X - E[X])] + E[f'_Y(\theta)(Y - E[Y])] \\ &= E[f(\theta)] + f'_X(\theta)E[(X - E[X])] + f'_Y(\theta)E[Y - E[Y])] \\ &= E[f(\theta)] \\ &\approx f(\theta) \\ &= \frac{E[X]}{E[g(Y)]} \\ &= 0. \end{aligned}$$

By definition, the variance of $f(X,Y)$ is given by

$$\text{Var}(f(X,Y)) = E\left[\left(f(X,Y) - E[f(X,Y)\right)^2\right]. \tag{24}$$

After substituting $E[f(X,Y)] \approx f(\theta)$, the first-order approximation of the variance is

$$\begin{aligned} \text{Var}(f(X,Y)) &\approx E\left[\left(f(X,Y) - f(\theta)\right)^2\right] \\ &\approx E\left[\left(f(\theta) + f'_X(\theta)(X - E[X]) + f'_Y(\theta)(Y - E[Y]) - f(\theta)\right)^2\right] \\ &= E[f'^2_X(\theta)(X - E[X]) + f'^2_Y(\theta)(Y - E[Y]) \\ &\quad + 2f'_X(\theta)(X - E[X])f'_Y(\theta)(Y - E[Y])] \\ &= f'^2_X(\theta)\text{Var}(X) + f'^2_Y(\theta)\text{Var}(Y) + 2f'_X(\theta)f'_Y(\theta)\text{Cov}(X,Y). \end{aligned}$$

Since $f'_X(X,Y) = \frac{1}{g(Y)}$, $f'_Y(X,Y) = \frac{-X}{g'^2(Y)}$, and $\theta = (0, E[Y])$,

$$\mathrm{Var}(f(X,Y)) \approx \frac{1}{g^2(E[Y])}\mathrm{Var}(X).$$

Q.E.D.

We can now state a proof of Proposition 4.

PROPOSITION 4 *For all uncoupled processes $Z_i(t)$, $i = 1 \ldots n$, with $\sigma(t) = \sigma$, and $E[t_i]$ which decreases in v_i, the resulting random variable $\tilde{\eta}_{ji}$ (via Proposition 3) has mean 0 and variance which increases in any element of* $\mathbf{v}$.

PROOF: Recall the expression for $\tilde{\eta}_{ji}$ from equation (21),

$$\tilde{\eta}_{ji} = \frac{\sigma \int_0^{t^*} U^{-1}(\tau)\, d(B_j(\tau) - B_i(\tau))}{\int_0^{t^*} U^{-1}(\tau)\, d\tau}.$$

Note that the stochastic process $B_j(\tau) - B_i(\tau)$ in the numerator is also a Brownian motion with independent increments, and the variance of an individual increment $d(B_j(\tau) - B_i(\tau))$ is 2. Since the integral in the numerator can be interpreted as the limit of sums (Smith, 2000), the variance of the numerator is given by $2\sigma^2 \int_0^{t^*} U^{-2}(\tau)\, d\tau$, by virtue of the fact that for a sequence of independent random variables, $\{X_n\}_1^n$, $\mathrm{Var}\Big(\sum_1^n f(n) X_n\Big) = \sum_1^n f^2(n)\mathrm{Var}(X_n)$.

Therefore (21) can be re-expressed in terms of a random variable $\tilde{B}$ with mean 0 and variance 1,

$$\begin{aligned}\tilde{\eta}_{ji} &= \frac{\sqrt{2\sigma^2 \int_0^{t^*} U^{-2}(\tau)\, d\tau}\tilde{B}}{\int_0^{t^*} U^{-1}(\tau)\, d\tau} \\ &= \frac{\sqrt{2}\sigma\tilde{B}}{C(t^*)},\end{aligned}$$

where

$$C(t^*) \equiv \frac{\int_0^{t^*} U^{-1}(\tau)\, d\tau}{\sqrt{\int_0^{t^*} U^{-2}(\tau)\, d\tau}}.$$

From Lemma 1, $C(t^*)$ is increasing in t^*.[20] Therefore we can apply Lemma 2 to $\tilde{\eta}_{ji}$, yielding

$$\text{(25)} \qquad \operatorname{Var}(\tilde{\eta}_{ji}) \approx \frac{2\sigma^2}{C^2(E[t^*])}.$$

For our result, we therefore must show that $E[t^*(\mathbf{v})]$ is decreasing in $\mathbf{v}$, with the dependence of t^* on the parameter $\mathbf{v}$ now explicitly stated.

A closed form for the distribution of $t^*(\mathbf{v})$ does not exist, even in the case of Brownian motion. However its moments can be calculated by considering $t^*(\mathbf{v}) = \min\{t_1, \dots, t_n\}$ as an order statistic with the following cumulative distribution function,

$$\text{(26)} \qquad H_{t^*}(t; \mathbf{v}) = 1 - \prod_{i=1}^{n} (1 - G_i(t; v_i)),$$

probability density function,

$$h_{t^*}(t; \mathbf{v}) = \sum_{i=1}^{n} g_i(t; v_i) \prod_{i=1, j \neq i}^{n} [1 - G_i(t; v_i)],$$

and mean

$$E[t^*(\mathbf{v})] = \int_0^\infty t h_{t*}(t; \mathbf{v})\, dt.$$

For any $\mathbf{v}'$ and $\mathbf{v}$ where $v'_j \geq v_j, \forall j$, and $\exists i$ where $v'_i > v_i$, then $E[t^*(\mathbf{v})] > E[t^*(\mathbf{v}')]$ if and only if

$$H_{t^*}(t; \mathbf{v}) < H_{t^*}(t; \mathbf{v}').$$

Since $E[t_i(v_i)]$ is decreasing in v_i, then $G_i(t; v_i) < G_i(t; v'_i)$ for all i and t, and (26) yields the result. *Q.E.D.*

---

[20] For the case of $U(t)$ decreasing in t (i.e. equation **??**), this result can be derived straight-forwardly from the Fundamental Theorem of Calculus.

7.4. *Proof of Proposition 5*

PROPOSITION 5 *For a Brownian motion accumulator (5) and a Race stopping rule (equation 9), then* $P_i(\mathbf{v}\boldsymbol{\alpha}^\top) > P_i(\mathbf{v})$ *for some* $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]$ *where* $\alpha_i > 1$ *and* $\alpha_j = 1, \forall j \neq i$.

PROOF: Recalling the choice probabilities given in equation 2, note that $\alpha_i v_i - v_j$ increases in order α_i. Therefore to ensure that

$$\begin{aligned} P_i(\mathbf{v}\boldsymbol{\alpha}^\top) &= \Pr\left[\alpha_i v_i - v_j > \tilde{\eta}_{ji}(\mathbf{v}\boldsymbol{\alpha}^\top), \quad \forall j \neq i\right] \\ &> \Pr\left[v_i - v_j > \tilde{\eta}_{ji}(\mathbf{v}), \quad \forall j \neq i\right] \\ &= P_i(\mathbf{v}) \end{aligned}$$

we must show that the standard deviation of $\tilde{\eta}_{ji}(\mathbf{v})$ increases at a slower rate, or equivalently, that the variance increases at a rate less than α_i^2. Let "Big-O" notation, $O_x(x^p)$, denote the order p of the quantity x^p as $x \to \infty$. Applying the order notation to the variance of $\tilde{\eta}_{ji}$ given in equation (25) yields

$$O_{\alpha_i}\Big(\mathrm{Var}\,(\tilde{\eta}_{ji})\Big) \approx O_{\alpha_i}\big(E^{-1}[t^*]\big).$$

Moreover, from equation (26),

$$O_{\alpha_i}\big(H_{t^*}(t;\mathbf{v})\big) = O_{\alpha_i}\big(G_{t_i}(t;\alpha_i v_i)\big).$$

Since α_i is a parameter in the stopping time distribution, it passes through the differentiation and subsequent integration required to get the densities and moments, therefore

$$O_{\alpha_i}(E[t^*]) = O_{\alpha_i}(E[t_i(\alpha_i v_i)]).$$

For a Brownian motion accumulator, $E[t_i(\alpha v_i)] = \frac{\theta}{\alpha v_i}$, therefore

$$O_{\alpha_i}(E[t^*]) = \alpha_i^{-1},$$

and we conclude,

$$O_{\alpha_i}\Big(\mathrm{Var}\left(\tilde{\eta}_{ji}\right)\Big) \approx O_{\alpha_i}\big(E^{-1}[t^*]\big) = \alpha_i.$$

Q.E.D.

REFERENCES

Batley, R. (2008): "On Ordinal Utility, Cardinal Utility and Random Utility," *Theory and Decision*, 64(1), 37–63.

Beck, J. M., W. J. Ma, R. Kiani, T. D. Hanks, A. K. Churchland, J. D. Roitman, M. N. Shadlen, P. E. Latham, and A. Pouget (2008): "Probabilistic population codes for Bayesian decision making," *Neuron*, 60(6), 1142–1152.

Becker, G. M., M. H. DeGroot, and J. Marschak (1963): "Stochastic models of choice behavior," *Behavioral Science*, 8(1), 41–55.

Bernheim, B. D. (2009): "On the potential of neuroeconomics: a critical (but hopeful) appraisal," *American Economic Journal: Microeconomics*, 1(2), 1–41.

Block, H. D., and J. Marschak (1960): "Random Orderings and Stochastic Theories of Responses," *Cowles Foundation Discussion Papers*.

Bogacz, R., E. Brown, J. Moehlis, and P. Holmes (2006): "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks.," *Psychological Review*, 113(4), 700–765.

Bogacz, R., M. Usher, J. Zhang, and J. L. McClelland (2007): "Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1485), 1655–1670.

Britten, K. H., M. N. Shadlen, W. T. Newsome, and J. A. Movshon (1992): "The analysis of visual motion: a comparison of neuronal and psychophysical performance.," *The Journal of Neuroscience*, 12(12), 4745–4765.

Buschena, D., and D. Zilberman (2000): "Generalized Expected Utility, Heteroscedastic Error, and Path Dependence in Risky Choice," *Journal of Risk and Uncertainty*, 20(1), 67–88.

Busemeyer, J. R., and J. Townsend (1992): "Fundamental derivations from decision field theory," *Mathematical Social Sciences*, 23(3), 255–282.

Camerer, C. F., and T.-H. Ho (1994): "Violations of the betweenness axiom and nonlinearity in probability," *Journal of Risk and Uncertainty*, 8(2), 167–196.

Churchland, A. K., R. Kiani, and M. N. Shadlen (2008): "Decision-making with multiple alternatives," *Nature Neuroscience*, 11(6), 693–702.

Cisek, P. (2006): "Integrated neural processes for defining potential actions and deciding between them: a computational model," *The Journal of Neuroscience*, 26(38), 9761–9770.

CLITHERO, J. A., AND A. RANGEL (2014): "Combining response times and choice data using a neuroeconomic model of the decision process improves out-of-sample predictions.," *Cal Tech Working Paper*.

COX, D. R., AND H. D. MILLER (1965): *The Theory of Stochastic Processes*. Chapman Hall Ltd.

DAGSVIK, J. K. (1995): "How Large is the Class of Generalized Extreme Value Random Utility Models?," 39(1), 90–98.

DITTERICH, J. (2010): "A Comparison between Mechanisms of Multi-Alternative Perceptual Decision Making: Ability to Explain Human Behavior, Predictions for Neurophysiology, and Relationship with Decision Theory.," *Frontiers in Neuroscience*, 4(184), 1–24.

DITTERICH, J., AND A. K. CHURCHLAND (2012): "New advances in understanding decisions among multiple alternatives," *Current Opinion in Neurobiology*, 22, 1–7.

DITTERICH, J., M. E. MAZUREK, J. D. ROITMAN, AND M. N. SHADLEN (2003): "A role for neural integrators in perceptual decision making," *Cerebral Cortex*, 13(11), 1257–1269.

DRUGOWITSCH, J., R. MORENO-BOTE, A. K. CHURCHLAND, M. N. SHADLEN, AND A. POUGET (2012): "The cost of accumulating evidence in perceptual decision making.," *The Journal of Neuroscience*, 32(11), 3612–3628.

DRUGOWITSCH, J., AND A. POUGET (2012): "Probabilistic vs. non-probabilistic approaches to the neurobiology of perceptual decision-making," *Current Opinion in Neurobiology*, 22(6), 963–969.

FALMAGNE, J. C. (1978): "A representation theorem for finite random scale systems," 18(1), 52–72.

FECHNER, G. (1860): *Elements of Psychophysics*, vol. 1. Holt, Rhinehart and Winston, New York.

FEHR, E., AND A. RANGEL (2011): "Neuroeconomic Foundations of Economic Choice—Recent Advances," *Journal of Economic Perspectives*, 25(4), 3–30.

GLIMCHER, P. W. (2011): *Foundations of Neuroeconomic Analysis*. Oxford University Press.

GLIMCHER, P. W., AND E. FEHR (2013): *Neuroeconomics*, Decision Making and the Brain. Academic Press, 2nd edn.

GOLD, J. I., AND M. N. SHADLEN (2002): "Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward.," *Neuron*, 36(2), 299–308.

——— (2007): "The neural basis of decision making," *Annual Review of Neuroscience*, 30, 535–574.

Gul, F., and W. Pesendorfer (2006): "Random Expected Utility," *Econometrica: Journal of the Econometric Society*, 74(1), 121–146.

Haile, P., A. Hortaçsu, and G. Kosenok (2008): "On the empirical content of quantal response equilibrium," *The American Economic Review*, 98(1), 180–200.

Harrison, G. W. (2008): "Neuroeconomics: A critical reconsideration," *Economics and Philosophy*, 24(03), 303–344.

Harrison, G. W., and E. E. Rutstrom (2008): "Risk Aversion in the Laboratory," in *Risk Aversion in Experiments*, ed. by J. C. Cox, and G. W. Harrison. Emerald Group Publishing Ltd.

Hausman, J., and D. A. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica: Journal of the Econometric Society*, 46(2), 403–426.

Heitz, R. P., and J. D. Schall (2012): "Neural mechanisms of speed-accuracy tradeoff," *Neuron*, 76(3), 616–628.

Hey, J. D. (1995): "Experimental investigations of errors in decision making under risk," *European Economic Review*, 39(3-4), 633–640.

——— (2005): "Why we should not be silent about noise," *Experimental Economics*, 8, 325–345.

Hey, J. D., and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica: Journal of the Econometric Society*, 62(6), 1291–1326.

Holt, C. A., and S. K. Laury (2002): "Risk aversion and incentive effects," *The American Economic Review*, 92(5).

Hunt, L. T., N. Kolling, A. Soltani, M. W. Woolrich, M. F. S. Rushworth, and T. E. J. Behrens (2012): "Mechanisms underlying cortical activity during value-guided choice.," *Nature Neuroscience*, 15(3), 470–6– S1–3.

Kiani, R., L. Corthell, and M. N. Shadlen (2014): "Choice Certainty Is Informed by Both Evidence and Decision Time," *Neuron*, 84(6), 1329–1342.

Kiani, R., T. D. Hanks, and M. N. Shadlen (2008): "Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment," *The Journal of Neuroscience*, 28(12), 3017–3029.

Kiani, R., and M. N. Shadlen (2009): "Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex," *Science*, 324(5928), 759–764.

Krajbich, I., C. Armel, and A. Rangel (2010): "Visual fixations and the computation and comparison of value in simple choice," *Nature Neuroscience*, 13(10), 1292–1298.

Krajbich, I., and A. Rangel (2011): "Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions.," *Proceedings of the National Academy of Sciences*, 108(33), 13852–13857.

Link, S. W., and R. A. Heath (1975): "A sequential theory of psychological discrimination," *Psychometrika*.

Liston, D. B., and L. S. Stone (2013): "Saccadic brightness decisions do not use a difference model," *Journal of Vision*, 13(8), 1–10.

Loomes, G. (2005): "Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data," *Experimental Economics*, 8(4), 301–323.

Loomes, G., and R. Sugden (1995): "Incorporating a stochastic element into decision theories," *European Economic Review*, pp. 641–648.

——— (1998): "Testing Different Stochastic Specificationsof Risky Choice," *Economica*, 65(260), 581–598.

Luce, R. D. (1986): *Response Times*. Oxford University Press, New York.

Luce, R. D., and P. Suppes (1965): "Preferences, utility and subjective probability," in *Handbook of Mathematical Psychology*, ed. by R. D. Luce, R. Bush, and E. Galanter, pp. 249–410. John Wiley and Sons, Inc.

Manski, C. F. (1977): "The structure of random utility models," *Theory and Decision*, 8(3), 229–254.

Marley, A. A. J., and H. Colonius (1992): "The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation," 36, 1–20.

Marschak, J. (1960): "Binary choice constraints on random utility indications," in *Stanford Symposium on Mathematical Methods in the Social Sciences*, ed. by K. Arrow. Stanford University Press.

Mazurek, M. E., J. D. Roitman, J. Ditterich, and M. N. Shadlen (2003): "A role for neural integrators in perceptual decision making.," *Cerebral Cortex*, 13(11), 1257–1269.

McFadden, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Economic Theory and Mathematical Economics*, ed. by P. Zarembka. Academic Press, Inc.

——— (1981): "Structural Discrete Probability Models Derived from Theories of Choice," in *Structural Analysis of Discrete Data and Econometric Applications*, ed.

by C. F. Manski, and D. L. McFadden. MIT Press.

——— (2001): "Economic choices," *The American Economic Review*, 91(3), 351–378.

——— (2005): "Revealed stochastic preference: a synthesis," *Economic Theory*, 26(2), 245–264.

McFadden, D. L., and K. Train (2000): "Mixed MNL models for discrete response," *Journal of Applied Econometrics*, 15, 447–470.

McMillen, T., and P. Holmes (2006): "The dynamics of choice among multiple alternatives," *Journal of Mathematical Psychology*, 50(1), 30–57.

Milosavljevic, M., J. Malmaud, A. Huth, C. Koch, and A. Rangel (2010): "The drift diffusion model can account for value-based choice response times under high and low time pressure," *Judgement and Decision Making*, 5(6), 437–449.

Mörters, P., and Y. Peres (2010): *Brownian Motion*. Cambridge University Press.

Niwa, M., and J. Ditterich (2008): "Perceptual Decisions between Multiple Directions of Visual Motion," *The Journal of Neuroscience*, 28(17), 4435–4445.

Ord, K., and A. Stewart (1994): *Kendall's Advanced Theory of Statistics*. 6th edn.

Ratcliff, R. (1978): "A theory of memory retrieval," *Psychological Review*, 85(2), 59–108.

Ratcliff, R., and P. L. Smith (2004): "A comparison of sequential sampling models for two-choice reaction time," *Psychological Review*, 111(2), 333–367.

Roitman, J. D., and M. N. Shadlen (2002): "Response of Neurons in the Lateral Intraparietal Area during a Combined Visual Discrimination Reaction Time Task," *The Journal of Neuroscience*.

Roxin, A., and A. Ledberg (2008): "Neurobiological Models of Two-Choice Decision Making Can Be Reduced to a One-Dimensional Nonlinear Diffusion Equation," *PLoS Computational Biology*, 4(3), e1000046.

Shadlen, M. N., T. D. Hanks, A. K. Churchland, and R. Kiani (2006): "The speed and accuracy of a simple perceptual decision: a mathematical primer," in *Bayesian Brain Probabilistic Approaches to Neural Coding*, ed. by K. Doya, S. Ishii, A. Pouget, and R. Rao.

Smith, P. L. (2000): "Stochastic Dynamic Models of Response Time and Accuracy: A Foundational Primer.," 44(3), 408–463.

Smith, P. L., and R. Ratcliff (2004): "Psychology and neurobiology of simple decisions.," *Trends in neurosciences*, 27(3), 161–168.

Thurstone, L. (1927): "A law of comparative judgment," *Psychological Review*, 34(4), 273–286.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edn.

Tsetsos, K., J. Gao, J. L. McClelland, and M. Usher (2012): "Using Time-Varying Evidence to Test Models of Decision Dynamics: Bounded Diffusion vs. the Leaky Competing Accumulator Model," *Frontiers in Neuroscience*, 6(79), 1–17.

Tsetsos, K., M. Usher, and N. Chater (2010): "Preference reversal in multiattribute choice," *Psychological Review*, 117(4), 1275–1293.

Usher, M., and J. L. McClelland (2001): "On the time course of perceptual choice: A model based on principles of neural computation," *Psychological Review*, 108(3), 550–592.

Wald, A., and J. Wolfowitz (1948): "Optimum Character of the Sequential Probability Ratio Test," *The Annals of Mathematical Statistics*, 19(3), 326–339.

Wang, X.-J. (2002): "Probabilistic Decision Making by Slow Reverberation in Cortical Circuits," *Neuron*, 36(5), 955–968.

Webb, R. (2011): "Does Neuroeconomics Have a Role in Explaining Choice?," *SSRN Working Paper*.

——— (2013): "Dynamic constraints on the distribution of stochastic choice: Drift Diffusion implies Random Utility ," *SSRN Electronic Journal*.

Webb, R., and M. C. Dorris (2013): "A Neural Model of Stochastic Choice in a Mixed Strategy Game," *SSRN Working Paper*.

Webb, R., P. W. Glimcher, I. Levy, S. Lazzaro, and R. B. Rutledge (2013): "Neural Random Utility and Measured Value," *SSRN Working Paper*.

Webb, R., P. W. Glimcher, and K. Louie (2014): "Rationalizing Context-Dependent Preferences: Divisive Normalization and Neurobiological Constraints on Decision-Making," *SSRN Working Paper*.

Weber, E. (1834): *On the Tactile Senses (with translation of De Tactu)*. Experimental Psychology Society, New York.

Wilcox, N. T. (2008): "Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison," in *Risk Aversion in Experiments*, ed. by G. W. Harrison, and J. C. Cox, pp. 197–292. Emerald Group Publishing Ltd.

Wong, K. F., and X.-J. Wang (2006): "A Recurrent Network Mechanism of Time Integration in Perceptual Decisions," *The Journal of Neuroscience*, 26(4), 1314–1328.

Woodford, M. (2014): "An Optimizing Neuroeconomic Model of Discrete Choice," *NBER Working Paper*.

Yellott Jr., J. I. (1977): "The relationship between Luce's Choice Axiom, Thurstone's

Theory of Comparative Judgment, and the double exponential distribution," 15(2), 109–144.