

Rationalizing Context-Dependent Preferences: Divisive Normalization and Neurobiological Constraints on Choice

Ryan Webb, Paul W. Glimcher, and Kenway Louie*

May 25, 2014

Abstract

Biology places a resource constraint on the form of neural computations that ultimately characterize stochastic choice behaviour. We incorporate one such computation, *divisive normalization*, into a neuroeconomic model and predict that the composition and size of the choice set will adversely influence choice. Evidence for novel violations of the IIA axiom is provided from two behavioural experiments, and observed behaviour is more accurately captured by a choice model incorporating normalization compared to existing econometric specifications. Finally, we demonstrate that normalization implements an efficient means for the brain to represent valuations given neurobiological constraints, yielding the fewest choice errors possible.

Whether stochastic choice behaviour can be rationalized, and how such behaviour depends on the choice set, has long been of interest in economics. Block and Marschak (1960) first laid out a necessary *regularity* condition for the existence of a random utility representation, stating that the probability of choosing any alternative must not increase as the choice set expands. Similarly, an early sufficient condition held that relative preferences (i.e. relative choice probabilities) between any two alternatives should be independent of the other elements of the choice set, formally stated as the axiom of Independence of Irrelevant Alternatives (IIA; Luce, 1959). However it was quickly recognized that IIA places strong conditions on choice behaviour (Debreu, 1960) and violations have

*Webb: Rotman School of Management, University of Toronto, Toronto, Ontario; Institute for the Interdisciplinary Study of Decision-Making, New York University, ryan.webb@utoronto.ca. Glimcher: Institute for the Interdisciplinary Study of Decision-Making, New York University, 4 Washington Pl., NY, NY, 10003, pg3@nyu.edu. Louie: Institute for the Interdisciplinary Study of Decision-Making, New York University, 4 Washington Pl., NY, NY, 10003, k1837@nyu.edu. The authors wish to thank A. Caplin, N.G. Martineau, T. Strzalecki, K. Train, and T. LoFaro for helpful comments, as well as seminar participants at Harvard University, the University of Zurich, and Maastricht University. We also thank M. Bukwich and M. Khaw for excellent research assistance. This research was funded by the Army Research Office (W911ND-11-1-0482), and the experiment was approved by an IRB.

been well-documented empirically at both the population and individual level (Rieskamp, Busemeyer and Mellers, 2006, for a review). These violations, compatible with some rational models, consist of patterns of substitution between alternatives as the composition of the choice set is varied. Therefore *how* relative preferences depend on the choice set is of great interest in both economic theory (McFadden, 2005; Karni and Vierø, 2013) and evaluating responses to economic policy in applications like transportation planning, energy usage, and consumer choice (McFadden, 1974; Hausman and Wise, 1978; Ben-Akiva and Lerman, 1985; Revelt and Train, 1998; Brownstone and Train, 1999; McFadden and Train, 2000; Train and Winston, 2007; Hess and Rose, 2009; see McFadden, 2001, for a review).

In this article, we examine novel forms of IIA violations that are necessarily predicted by a resource constraint known to be imposed on neurobiological systems. The observation that biology influences choice behaviour – and should explain economic characteristics – is not novel in the economic literature (Robson, 2001; Robson and Samuelson, 2010; Netzer, 2009). This includes the observation that a constrained decision process implies behaviour we might strictly term irrational ex-ante, but is in fact an optimal response to some biological constraint (Simon, 1979). The recent literature on *rational inattention*, for example, examines how constraints on information-processing (Shannon, 1948) can explain observed departures from rational expectations, expected utility, and the traditional random utility model (Sims, 2003; Woodford, 2012; Caplin and Dean, 2014). Similarly, Rayo and Becker (2007) demonstrate that the utility representation must be adaptive and depend on the distribution of the domain of utility, given constraints on the representation similar to those we identify. Here, our contribution is to provide a neural mechanism for how such choice-set dependence arises from neurobiological constraints, explore its predictions for substitution patterns between choice alternatives, provide behavioural evidence for its existence, and demonstrate that it implements a solution to the neurobiological choice problem.

Like constraints traditionally described in economics, neurobiological constraints arise from scarcity: neurons require energy in order to transmit information. Energy resources in the brain must be allocated to various neural systems for various tasks, and there is some mechanism, presumed to be evolution, which selects more efficient allocations and rejects less efficient ones. The evolution of the decision-making process in the brain is directly shaped by the scarcity of resources. In particular, there is a limit on both the total number of neurons and the number that can be allocated to any given task (since neurons are not perfectly substitutable). This basic capacity constraint has two implications that are particularly relevant for modelling choice behaviour, which we briefly preview.

Neural activity is universally observed to be stochastic. Stochasticity in neural activity is, at least in part, due to the small-scale, thermodynamic processes involved in the transmission of information between neurons. While

this randomness can be partially mitigated through increasing the number (or size) of neurons that participate in any task, this is costly in terms of resources. Moreover there may be a limit to the effectiveness of pooling over a large number of neurons: for the purposes of redundancy and computational speed, neurons share inputs and are not fully independent in their activity (Shadlen and Kiani, 2013, Box 1). Ultimately, this neural stochasticity results in stochastic choice behaviour (Glimcher, 2005) which has been shown to arise optimally under a capacity constraint on information processing (Woodford, 2014). For our purposes, neural stochasticity can be related to choice behaviour in a familiar and flexible econometric framework called the Neural Random Utility Model (NRUM; Webb et al., 2013), which we will find useful for capturing the second implication of a neurobiological resource constraint.

The brain encodes information in relative terms. The relative coding of information by neural systems is one of the fundamental observations of neuroscience in the past century (Hartline and Ratliff, 1957). Because neural activity is bounded, in terms of the dynamic range of a single neuron, the finite number of neurons in a brain area, and finite noise, evolutionary forces have implemented computational algorithms which compress the objective state of the world for subjective perception, and ultimately, choice. There is a long literature in neuroscience examining the structure and operation of neural computations, and more recently, neuroeconomists have been studying them in the context of choice behaviour (Glimcher, 2011; Fehr and Rangel, 2011). In this article, we focus on a pervasive computation in cortex called *divisive normalization* (Heeger, 1992; Schwartz and Simoncelli, 2001; Carandini and Heeger, 2012) which implements a relative coding of information and has recently been observed in value signals that are believed to participate in the implementation of a decision (Louie, Grattan and Glimcher, 2011).

In principle, a neural computation is implemented by a collection of neurons that are connected together in a systematic manner. In the abstract, these computations are typically thought of as mathematical functions with inputs and outputs. The divisive normalization computation can be represented by the following (simplified) functional form,

$$f(\mathbf{x}) = \frac{\mathbf{x}}{(\sum_{n \in N} x_n^\beta)^{\frac{1}{\beta}}}, \quad (1)$$

in which an input vector $\mathbf{x} = [x_1, \dots, x_N] \in \mathbb{R}_+^N$ is scaled by the magnitude and number of its elements in the form of a norm of degree β .¹ This computation normalizes each input with respect to the entire set of inputs, yielding a relative relationship between them. It is this relative relationship that we will explore in the context of choice behaviour, interpreting the input \mathbf{x} as a decision maker's valuation of a choice alternative in a choice set of size N .

¹A norm of order p is defined as $\|x\|_p \equiv (\sum_{n=1}^N |x_n|^p)^{1/p}$ where $p > 0$.

Originally observed in the cortical regions of the brain involved in visual perception (Heeger, 1992), divisive normalization has now been observed in multiple forms of sensory and value processing and across species ranging from invertebrates to primates (Louie, Grattan and Glimcher, 2011; Carandini and Heeger, 2012). From a normative standpoint, divisive normalization has also been shown to yield an efficient coding of sensory information in a constrained neural system (Schwartz and Simoncelli, 2001; Wainwright, Schwartz and Simoncelli, 2001; Sinz and Bethge, 2013; Qamar et al., 2013). However, the normative implications for behaviour have remained unclear, particularly because divisive normalization predicts context-dependent choice behaviour that, in the absence of constraints, can be strictly termed inefficient and inconsistent (Louie, Khaw and Glimcher, 2013).

To establish both the positive and normative implications of normalization in the domain of choice, we extend an NRUM (Webb et al., 2013) to directly incorporate the divisive normalization computation. The extended model predicts how relative choice probabilities depend on the size and composition of the choice set, and yields novel substitution patterns between alternatives as the choice set is varied. Of particular note, the model predicts a decrease in the likelihood of choosing a higher-valued alternative as the choice set is expanded or as low-valued alternatives are manipulated. Evidence for such context-dependent choice behaviour is provided from two behavioural experiments: one which varies the composition of the choice set and another which varies its size, and the substitution patterns we observe are more accurately captured by an NRUM incorporating neurobiological constraints than common econometric specifications. Finally, we show that divisive normalization yields such behaviour by implementing a solution to the decision-making problem faced by a resource-constrained neural system.

From an empirical standpoint, there is now growing evidence that economic behaviour depends systematically on the choice set. Of particular relevance to our study, the size of the choice set has been shown to alter relative preference, the consistency of preferences, and the likelihood of making a purchase (Iyengar and Lepper, 2000; DeShazo and Fermo, 2002). The model we present here is able to capture such effects by focusing on normalization solely at the level of choice alternatives. This is in contrast to existing models which capture systematic variation along the “attribute” dimension of alternatives such as colour, size, or quality. For instance, it is often observed that adding an alternative to a choice set which is either dominated (or similar) in some attribute can alter relative choice towards one alternative — an empirical result referred to as *asymmetric dominance* (or *attractor* effects) (Huber, Payne and Puto, 1982; Tversky and Simonson, 1993; Rieskamp, Busemeyer and Mellers, 2006; Soltani, De Martino and Camerer, 2012).² Though the model we present maintains an ordinal ranking over alternatives, an extension of relative coding to the attribute dimension could capture such phenomena.

²Theoretical explanations of these “decoy” effects, and of the general relationship between context and choice, have been proposed (Thaler, 1985; Busemeyer and Townsend, 1993; Thaler, 1999; Bordalo, Gennaioli and Shleifer, 2012, 2013).

This article proceeds as follows. In section 1 we discuss the neurobiological resource constraint in more detail and present a general empirical framework for decision-making in the brain that will allow us to explore the implications of this constraint. We then introduce the divisive normalization computation, review its positive and normative application in perceptual neuroscience, and review the neural evidence for this representation in areas of the brain known to implement decision-making. In section 2, we explore the positive predictions of divisive normalization, specifically patterns in choice behaviour that depend on the composition and size of the choice set. We then present evidence for normalization from two behavioural experiments and propose a method for incorporating normalization into an econometric specification.³ Finally, in section 4 we address the normative issue by demonstrating that the divisive normalization computation implements choice behaviour that is optimal given the neurobiological resource constraint we identify. These results emphasize both the positive advances offered by developing choice models grounded in neuroscience, and the normative role neurobiological constraints can play in the study of decision-making.

1 Neurobiological Constraints on Decision-Making

We consider a neural system which is tasked with representing the value of each alternative in a choice set for the purpose of identifying the highest-valued alternative.⁴ We assume that evolutionary processes have allocated a neural resource of bounded size to this system, leaving open only the question of how decision-making is implemented given this bound. This assumption rests on the empirical observation that the decision-maker has limited endogenous control over the resources allocated to decision-making. Some background on the structure of the brain might be helpful in justifying this approach.

The total pool of neurons that make up the cortex of the brain can be anatomically divided into non-overlapping subsets of neurons, and each of these “areas” is distinct in structure and organization. Empirical evidence suggests that any given area serves a limited set of functional roles – which are the same across individuals – and a small subset of these areas have been identified as having a role in decision-making. It is essentially required by this logic that devoting more neural resources to these decision-making areas must balance the benefits of improved decision-making against the costs of that allocation to other functions, such as processing incoming sensory information or managing the heartbeat.⁵

³Results from one of these experiments, which varies the composition of the choice set, have previously been reported in the neuroscience literature in a related context (Louie, Khaw and Glimcher, 2013). In this article, we bring our econometric specification to this dataset in order to compare it to established econometric methods, investigate subject-specific predictions, and establish its normative implications.

⁴We will define “value” more explicitly shortly. For now, we just consider it a variable which is compared and yields choice behaviour.

⁵Of course, a full normative explanation of the resources allocated to decision-making

So why is a decision-maker limited in its ability to recruit more neural resources, even if it might be momentarily advantageous? The reason is that brain areas are distinct in their connections with the rest of the brain, distinct in their cellular structure, and distinct in many of their biophysical properties. These specializations, which distinguish different brain areas and make them appropriate for particular functions, develop over years – primarily over the first 20 years of development – and cannot be altered quickly. This imposes a significant constraint on the ability of the decision-maker to endogenously control the allocation of neural resources to a task. For these reasons, we take the allocation of neural resources to be effectively fixed at a particular moment in time, at least for the time horizon in which decisions are typically made by an individual.⁶ As we alluded to earlier, this constraint has two neurobiological implications that we aim to incorporate in a choice model. We now describe these implications in more detail.

1.1 Neural Activity is Stochastic

A characteristic feature of neural systems is an inherent stochasticity in the generation of neural activity. For example, repeated activation of a sensory neuron with the same stimulus will yield considerable variability in the magnitude and temporal pattern of activity elicited. This stochasticity is believed to arise from the biophysical mechanisms by which neurons transmit information. Neurons propagate information using discrete, all-or-none changes in their membrane voltage termed *action potentials*, and this neural activity is typically quantified by the temporal frequency of such events (e.g. action potentials per second). Communication occurs at a small gap between neurons (the *synapse*), where the electrical activity of the upstream (or pre-synaptic) neuron is converted into the release of molecules (*neurotransmitters*), which diffuse across the synaptic space and are subsequently reconverted into electrical activity in the downstream (or post-synaptic) neuron.

Stochasticity in this system largely arises from the thermodynamic processes governing molecular interactions at the synapse, for example the amount of neurotransmitter release and its effect on post-synaptic activity. The probability of a single pre-synaptic action potential resulting in a post-synaptic action potential is actually quite low and variable, and susceptible to randomness at the molecular level (Ribault, Sekimoto and Triller, 2011). While these issues can be partially resolved by increasing the size of neurons (thereby increasing the number of molecules involved in these mechanisms), or by increasing the number

would require a complete accounting of the costs of decision-related errors to the chooser and the metabolic costs of relaxing the constraints we identify — on an evolutionary timescale.

⁶There is a large literature in neuroscience and psychology which examines the prevalence of errors as a function of the time spent on a decision, known as the *speed/accuracy tradeoff* (for a review, see Bogacz et al., 2010). This literature finds that one can improve accuracy with longer decision times, but that this increase in accuracy is bounded. The fact that increasing the time devoted to a decision cannot overcome this resource limitation problem reflects a temporal constraint that has been both widely observed and biologically explained, though is not the subject of this article.

of neurons and pooling over more of them, this would require more resources. We also note that the nature of neural computation limits the stochasticity that can be reduced by pooling, for instance neurons receive correlated inputs for issues of redundancy (Shadlen and Newsome, 1998).

We capture neural stochasticity within a choice model by extending the Random Utility Model (RUM; Becker, DeGroot and Marschak, 1963; McFadden, 1974; Falmagne, 1978) to neural systems. The Neural Random Utility Model (Webb et al., 2013) is a formulation of the traditional RUM for stochastic choice within an individual, but one in which the unobservable latent variables are replaced with observable neural quantities.⁷ As such, the NRUM attempts to capture features of decision-making at the neural level within a familiar economic framework, and can be shown to be a reduced form of more structural, dynamic models found in neuroscience (Webb, 2013).

The basic formulation of an NRUM is based on over fifteen years of empirical literature which points towards a basic “two-stage” working hypothesis for decision-making in the primate brain (Glimcher, 2011; Fehr and Rangel, 2011). As an organizing principle, we incorporate these two stages explicitly in the NRUM by defining the observable neural quantity we call *subjective value*. Subjective value is a measurement of neural activity in the frontal cortex (and basal ganglia) of the human brain that carries information about the relative valuations of choice alternatives in a cardinal manner (Webb et al., 2013). Three recent meta-studies (Levy and Glimcher, 2011; Bartra, McGuire and Kable, 2013; Clithero and Rangel, 2013) now unambiguously indicate that this neural activity, particularly observed in the medial prefrontal cortex, is tightly correlated with every known economic method for estimating the values subjects place on choice alternatives. This includes revealed preferences over attributes (Tusche, Bode and Haynes, 2010; Levy et al., 2011; Smith et al., 2013), risk and ambiguity (Levy and Glimcher, 2012; Levy et al., 2010), and time (Kable and Glimcher, 2007, 2010). In our modelling, we denote the subjective value of each alternative in a choice set N by the vector $\mathbf{v} = [v_1, \dots, v_N] \in \mathbb{R}_+^{\|N\|}$, with the valuation of each alternative i denoted v_i . Since subjective value is encoded by stochastic neural activity, it is a random variable with density $g(\mathbf{v})$.

Once subjective values \mathbf{v} are instantiated in the activity of neurons, they must be compared for a choice to be executed. This comparison constitutes the second stage of the NRUM. The electrophysiological evidence suggests that (at least in the choice environments studied to date) regions of the frontal and/or parietal cortex receive subjective valuations as input and determine which of them is largest by means of a competition between neural quantities (Sugrue, Corrado and Newsome, 2004; Padoa-Schioppa and Assad, 2006; Gold and Shadlen, 2007; Louie and Glimcher, 2010, but also see Padoa-Schioppa 2011).⁸ Importantly, this competition is both stochastic and dynamic (Gold and

⁷We discuss the relationship between the NRUM and the RUM in more detail in section 2, and in Webb et al. (2013).

⁸For instance, in the Lateral Intra-Parietal area (LIP) – a neural circuit in monkeys that guides the selection of eye movements (or saccades) – groups of neurons each represent a saccade to a particular region of visual space. A decision between two saccadic options is

Shadlen, 2007; Kiani, Hanks and Shadlen, 2008; Milosavljevic et al., 2010), and structural models of this process (e.g. the drift diffusion model (Ratcliff, 1978)) can be captured by an additive noise vector we denote $\boldsymbol{\eta} = [\eta_1, \dots, \eta_N] \in \mathbb{R}^{\|\mathcal{N}\|}$ (Webb, 2013). For now, we will assume a general density $f(\boldsymbol{\eta})$.

The NRUM is then formally defined by a decision vector $\mathbf{u} = [u_1, \dots, u_N]$ comprised of subjective value and its stochastic comparison. For example, the model can be instantiated in the familiar additive form $\mathbf{u} = \mathbf{v} + \boldsymbol{\eta}$, with the choice of alternative i corresponding to $u_i > u_j, \forall i \neq j$. Figure 1 depicts this relation graphically for a binary choice set. The stochasticity in the comparison between v_1 and v_2 is depicted by the shaded error density, therefore the location of the vector \mathbf{v} relative to the 45° line determines the probability that the alternative with the highest value is chosen.

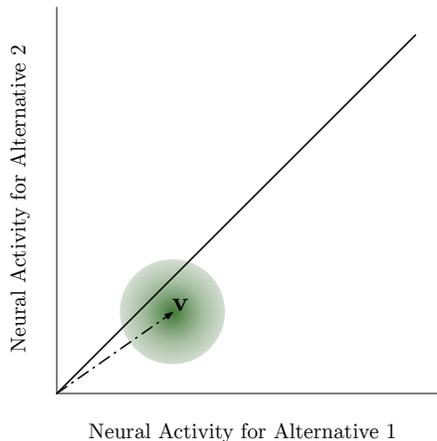


Figure 1: The basic NRUM for a binary choice.

Because the NRUM is composed entirely of neural quantities, it provides a convenient empirical framework for exploring the implications of neurobiological constraints on choice behaviour. In particular, it will allow us to capture the possibility that valuations are encoded relative to the other alternatives. To set the stage, let us define the transformation $\mathbf{z} = [z_1, \dots, z_N] = Z(\mathbf{v}) : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ with the resulting distribution $h(\mathbf{z})$. In general terms, the function $Z(\cdot)$ allows the transformed valuation of an alternative, z_i , to depend on the entire vector of valuations for all alternatives \mathbf{v} . Shortly, we will explore the implications when $Z(\mathbf{v})$ is defined to be the divisive normalization equation.

instantiated as a competition between two groups of neurons representing the two possible eye movements (Roitman and Shadlen, 2002; Kiani, Hanks and Shadlen, 2008). In the LIP, value modulation biases the competition towards the better option implementing a value-guided decision. Like frontal cortical areas, activity recorded in these areas varies with both the magnitude and probability of reward associated with specific actions in lotteries (Platt and Glimcher, 1999), in strategic games (Dorris and Glimcher, 2004), and with discounting over time delays (Louie and Glimcher, 2010).

All together, this yields a familiar framework. The decision vector is

$$\mathbf{u} = Z(\mathbf{v}) + \boldsymbol{\eta}, \tag{2}$$

and the decision-maker chooses option i when

$$\begin{aligned} u_i &> u_j \quad \forall j \neq i \\ z_i + \eta_i &> z_j + \eta_j \quad \forall j \neq i \end{aligned}$$

yielding a probability of choosing i

$$\begin{aligned} P_i &= P(z_i - z_j > \eta_j - \eta_i \quad \forall j \neq i) \\ &= \int \int \mathbb{1}[z_i - z_j > \eta_j - \eta_i, \quad \forall j \neq i] h(\mathbf{z}) f(\boldsymbol{\eta}) dv d\boldsymbol{\eta} \end{aligned} \tag{3}$$

assuming that \mathbf{v} (therefore \mathbf{z}) and $\boldsymbol{\eta}$ are independent, with $\mathbb{1}[\cdot]$ indicating the statement in brackets holds.

To isolate the behavioural implications of normalization, we make two simplifying assumptions. First, we only model normalization that takes place between brain regions which encode subjective value and brain regions which compare subjective values, as represented by the function $Z(\cdot)$. We strongly suspect that normalization occurs in other stages of the decision-making process, for instance in the construction of subjective value, \mathbf{v} , from objective sensory attributes, but do not model it here.⁹ Second, as a simplification we assume no stochasticity in subjective value \mathbf{v} , so the density $h(\mathbf{z})$ is degenerate and the probability of choosing i is given by

$$P_i = \int \mathbb{1}[z_i - z_j > \eta_j - \eta_i, \quad \forall j \neq i] f(\boldsymbol{\eta}) d\boldsymbol{\eta}. \tag{4}$$

In essence, we are assuming that we observe a particular realization of \mathbf{v} , and are investigating the impact of comparing alternatives in a choice set by means of the normalization $Z(\cdot)$ and the stochastic process captured by $\boldsymbol{\eta}$.

1.2 Neural Activity is Bounded

The second implication of our neurobiological resource constraint is that neural activity is bounded, both in terms of the activity range of a single neuron and the finite number of neurons in a brain area devoted to decision-making. As mentioned previously, we take the size of the brain area, and the number of neurons allocated to the representation of value, as fixed. We also note the existence of a maximal firing rate for all neurons due to the biophysical mechanism underlying the generation of neural activity: the individual events (*action*

⁹In fact, the existence of some form of normalization in these regions is a necessity considering the neural constraints we will highlight below. This raises the possibility that attributes may be independently normalized and would be sufficient for generating behaviour that violates a stable ordinal ranking of alternatives, such as the asymmetric dominance effect (Bordalo, Gennaioli and Shleifer, 2013).

potentials, or *spikes*) that constitute a neuron’s activity cannot repeat within a minimum amount of time (*refractory period*; Hodgkin and Huxley, 1952). All together, this requires the brain to implement computational algorithms which compress information into this bounded range with direct consequences for both sensory processing and choice behaviour.

We focus on a computation known as *divisive normalization* which implements a relative coding of information in response to this bound constraint. To understand how divisive normalization influences choice behaviour, it will first be useful to understand its origins in the sensory neuroscience literature. We then review the evidence for divisive normalization in value systems and demonstrate how it implements a bound constraint.

1.2.1 Divisive Normalization in Sensory Systems

A fundamental goal in *systems neuroscience* is to understand how the brain processes and represents information at the level of individual neurons. Much of this work has focused on the sensory domain in which an environmental stimulus is transduced through the sensory organs and passes through a series of processing stages (i.e. spatially and functionally distinct networks of neurons), each of which has a particular role. In the visual system, for instance, photons (which comprise light) make initial contact with photoreceptors in the retina and this information passes through subsequent stages of hierarchical neural processing in retinal ganglion cells, the thalamus, primary visual cortex, and so on. Generally speaking, sensory neurons in these regions are “tuned” for a particular stimulus, for instance the intensity of light in a particular region of visual space, or a particular direction of motion.

At each of these levels, initial theories of neural processing hypothesized that sensory neurons perform a linear computation: the activity of a neuron is simply a weighted sum of the inputs to that neuron. However, it was soon discovered that sensory neurons display activity that cannot be solely explained by linear models. One prominent non-linearity is the saturation of neural activity at high input levels. A second is the phenomenon of *suppression*, where the neural activity elicited by a stimulus (for which the neuron is tuned) is reduced in the presence of a stimulus for which it is not tuned.

The divisive normalization computation was proposed to explain such non-linear phenomena in primary visual cortex (Heeger, 1992). The critical feature of this model is a divisive rescaling of the activity of a population of neurons by the total activity of a larger pool of neurons. The normalization model can be defined, in its most general form, by the function

$$f_i(x_1, \dots, x_N) = \kappa \frac{x_i^\alpha}{\sigma + (\sum_n \omega_n x_n^\beta)^{\frac{1}{\gamma}}}, \quad (5)$$

where the response $f_i(x_1, \dots, x_N)$ of a population $i \in N$ is a function of both the driving input x_i to that population and the inputs to the other elements of the normalization pool N . This general response function is governed by a number

of parameters: κ denotes the maximum possible activity level, σ determines how responses saturate with increased driving input, ω_n weights each neural population in the normalization pool, and α , β and γ mediate the form of the normalization.

The divisive nature of the normalization implements several features consistent with how neurons in visual cortex respond to sensory information. First, because the input to a given neuron is also included in the normalization pool, the model produces a saturating response to increases in driving input. This is consistent with the observation that neurons are biologically constrained below a maximum firing rate. Second, division expresses the quantity encoded by a population of neurons in relative terms, scaled to the pooled activity of other inputs. This relative representation of information is observed to be a general feature of sensory coding. In the retina, for example, light intensity is encoded relative to the average ambient illumination; this process of light adaptation is why black type on white newspaper appears the same in a darkened room and under bright sunlight, even though the newsprint may reflect a million times more photons in the latter environment.

The ubiquity of the normalization computation suggests that it may be a general feature of information processing in cortical systems (for a review, see Carandini and Heeger, 2012). Originally proposed to explain non-linearities in primary visual cortex, normalization has been identified in multiple stages of visual processing from the retina to downstream visual areas, and in species ranging from the fruit fly to primates. In addition to vision, normalization characterizes neural responses in other sensory modalities such as audition and olfaction, and extends to higher-order cognitive processes including attention and multi-sensory integration. However, we should note that the normalization model describes a computational algorithm rather than a specific biophysical implementation. Researchers have identified a number of potential neural mechanisms for implementing normalization and it is likely that the normalization computation is mediated by different processes in different species and systems. Moreover, since evolution selects on the fitness of the output of a particular neural system, the parametrization of the normalization equation will depend on the role of the particular system and statistical features of its input. The widespread occurrence and varying implementations indicate that it is the normalization computation, rather than the specific mechanism and/or parameterization, which is critical to efficient neural systems.

Normative Implications in Sensory Systems Given the ubiquity of divisive normalization in sensory systems, the question of its normative role in neural coding has been raised. In systems neuroscience, the benchmark normative framework for information processing is the *efficient coding* hypothesis (Barlow, 1961) based on Shannon’s seminal work on information transmission via finite capacity channels. Because neural systems face necessary biological constraints (e.g. maximum firing rates and/or numbers of neurons), this hypothesis holds that sensory stimuli should be represented in a manner that minimizes

the redundancy in encoded information; that is, neural responses should be statistically independent. Crucially, the statistics of the natural environment are decidedly not independent. For instance, regions of visual space with intense light occur in clusters, unlike the statistically independent “snow” displayed on an un-tuned analog television set. Under the efficient coding hypothesis, sensory systems should transform information about the natural world into a less redundant representation by incorporating these dependencies directly in coding algorithms.

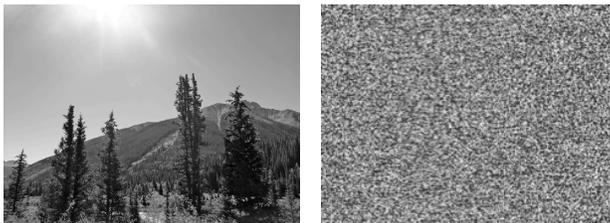


Figure 2: In the natural environment, nearby regions of visual space tend to be correlated (left), not independent (right).

While linear weighting functions can remove some redundancies, the statistics of natural images are too complex for linear models to produce completely independent responses. However, the non-linear divisive normalization model markedly reduces higher-order correlations in responses to both natural images and sounds (Schwartz and Simoncelli, 2001; Lyu, 2011; Sinz and Bethge, 2013), and has been shown to implement near-optimal categorization and encoding of these sensory stimuli (Qamar et al., 2013). Thus there is compelling evidence that normalization serves a specific normative role in implementing efficient information coding in sensory systems.

1.2.2 Divisive Normalization in Value Systems

Given the pervasive observation of divisive normalization in sensory brain areas, and the general similarity of information processing in neural systems, this has raised the question of whether a form of normalization is also taking place in regions of cerebral cortex that are involved in decision-making. Recent neural evidence from primates suggests that this is the case (Rorie et al., 2010; Louie, Grattan and Glimcher, 2011; Pastor-Bernier and Cisek, 2011). For example, Louie, Grattan and Glimcher (2011) examined the activity of neurons in the monkey lateral intraparietal area (LIP), a brain region involved in the selection of actions. Individual LIP neurons are active for specific eye movements; the activity of a given LIP neuron varies with the quantity of the reward associated with action completion (Platt and Glimcher, 1999). In the Louie, Grattan and Glimcher (2011) experiment, multiple alternatives were displayed on a screen, and the value of one alternative (in ml of water) was held constant

while the number and value of the other alternatives were varied. In the absence of normalization, varying the other alternatives should elicit no change in activity from the neuron which encoded for the constant target. Nonetheless, the measured neural activity varied inversely with the number and value of the alternatives, consistent with a model in which the neural activity coding the value of a specific alternative is suppressed by the value of others (Figure 3).

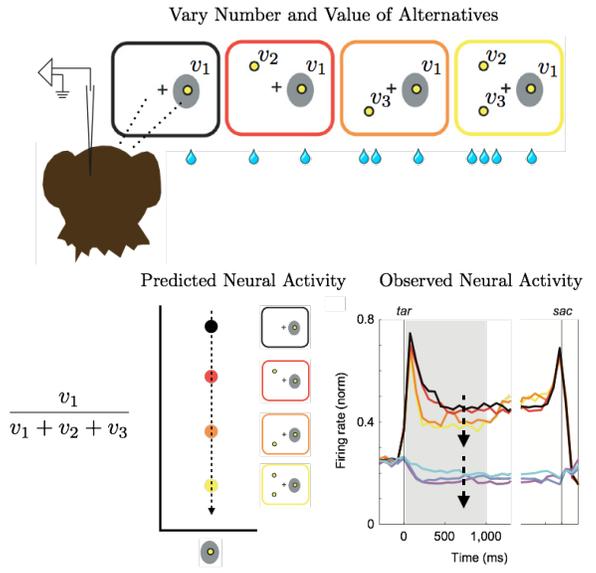


Figure 3: The Louie, Grattan and Glimcher (2011) experiment in which a monkey is evaluating choice alternatives. Neural activity is recorded for a constant alternative v_1 (grey circle), while the number and value (in μl of water) of the other alternatives is varied. The average activity of the recorded neurons over the course of the trial is reported.

To test this hypothesis, Louie, Grattan and Glimcher (2011) demonstrated that the following simplified normalization function,

$$\text{neural activity}_i = \kappa \frac{(\text{ml of water})_i}{\sigma + \sum_n (\text{ml of water})_n}, \quad (6)$$

captured the activity of LIP neurons better than either a simple linear form or other potential models (e.g. the difference in value).

This observation of normalization in value regions has implications for both both positive and normative theories of choice behaviour because it transforms neural activity and produces an output that is believed to drive choice. Since the value of all alternatives in a choice set are scaled by an identical term, ordinality is preserved, however the precise activity level corresponding to a given alternative depends on the other alternatives in the choice set. Given

variability in value representations, the interaction of scaling and stochasticity can have interesting implications for choice behaviour. The prevalence of divisive normalization in the brain also suggests a normative role related to efficient coding in the face of neurobiological constraints, but we will hold this question for section 4 and begin with the positive predictions.

1.2.3 Incorporating the Bound Constraint via Normalization

In section 1.1, we introduced the function $Z(\mathbf{v}) = [z_1, \dots, z_N]$ to capture a potential relationship between valuations of alternatives when they are compared. We now impose a bound \check{z} on the neural activity allocated to representing each z_i , and demonstrate that this bound constraint is implemented by defining $Z(\mathbf{v})$ to be the divisive normalization computation.

Figure 4 depicts three possible versions of a bound constraint on neural activity for a two-alternative choice set. Since increasing neural activity up to \check{z} requires resources, it is possible that there is some relationship or tradeoff between the neurons encoding the valuations of different alternatives. This relationship can be captured by a constraint in \mathbb{R}_+^N . At its least restrictive, the neural activity of each alternative is non-rivalrous and any activity levels interior to the square in Figure 4 are feasible. At the other extreme, the firing rates summed over alternatives could be bounded ($\sum_i z_i \leq \check{z}$), hence perfectly rivalrous, resulting in a linear constraint. Of course, any non-linear relationship between these alternatives can also be captured by a constraint that lies between these two extremes.

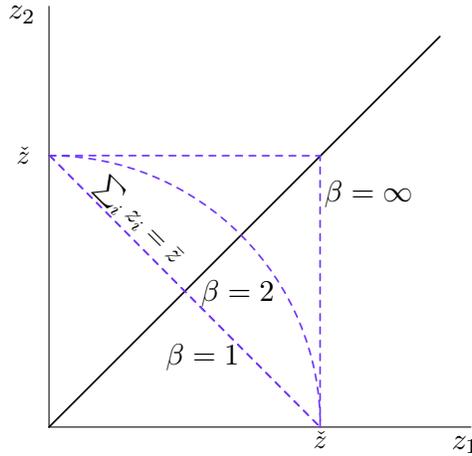


Figure 4: Three possible neural resource constraints that capture a tradeoff in neural activity.

The critical feature of the divisive normalization computation is that it scales neural activity to the boundary of a constraint, depending on the value of β .

We will work with a restricted version of the general normalization equation (5)

$$z_i = \kappa \frac{v_i}{\sigma + \omega \sum_{n \in N} (v_n^\beta)^{\frac{1}{\beta}}} \quad (7)$$

in which each neural population is weighted equally ($\omega_n = \omega$) and the denominator is restricted to be a norm of degree β . To aid exposition, we will also present the model under the assumption that $\sigma = 0$. Appendix D explores the implications of relaxing the assumption on σ in more detail, and the parameter will be unrestricted in our empirical results.

Let us begin with the linear constraint. If $\beta = 1$, then the normalization function $Z(\mathbf{v})$ scales the vector \mathbf{v} along a ray to the origin, proportioning neural resources between the alternatives along the line:

$$\sum_{i \in N} z_i = \sum_{i \in N} \kappa \frac{v_i}{\omega \sum_{n \in N} v_n} = \frac{\kappa \sum_{i \in N} v_i}{\omega \sum_{n \in N} v_n} = \kappa/\omega. \quad (8)$$

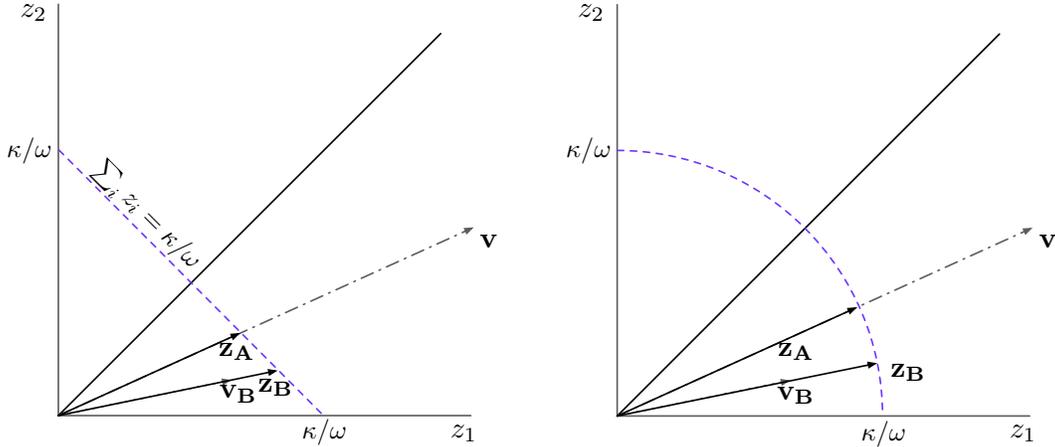


Figure 5: The proportionate scaling implemented by normalization for $\beta = 1$ (left) and $\beta = 2$ (right).

We now can reinterpret the ratio κ/ω as implementing the constraint \check{z} (Figure 5), and this relation emphasizes the role of ω in scaling the model. If $\omega = 0$ then the normalization term disappears, and along with it the means to bound \mathbf{z} .

For any $\beta > 0$, divisive normalization maps \mathbf{v} to a general class of geometric figures known as a *super-ellipse*. For instance, when $\beta = 2$, normalization scales the vector \mathbf{v} to the quarter-circle of radius κ/ω , since $\mathbf{z} = \frac{\kappa}{\omega} \frac{\mathbf{v}}{\|\mathbf{v}\|}$. Setting $\beta = 3$ maps \mathbf{v} to a hyperellipse, while $\beta = \infty$ maps to the square of size κ/ω . Notice that in each of these cases, divisive normalization preserves the relative magnitudes of the elements of \mathbf{v} , and scales neural activity to the boundary constraint

defined by \check{z} . Therefore divisive normalization implements a relative coding of value, subject to the boundary constraint on neural activity. In the following section, we will explore the implications of this result for choice behaviour.

2 Behavioural Implications of Normalization

The normalization of subjective valuations yields a number of interesting predictions for how choice behaviour depends on the choice set. These predictions stem from the interaction between the normalization equation (7) and the stochasticity present in the NRUM. To explore these predictions formally, we incorporate equation (7) directly in the NRUM defined in section 1.1. This yields a conditional probability of choosing alternative i given by:

$$P_i(\mathbf{v}) = \int \mathbb{1} \left[\kappa \frac{v_i - v_j}{\sigma + \omega (\sum_n v_n^\beta)^{\frac{1}{\beta}}} > \eta_j - \eta_i, \quad \forall j \neq i \right] f(\boldsymbol{\eta}) d\boldsymbol{\eta}. \quad (9)$$

In this section, we explore the behavioural predictions predicted by this equation as the choice set is varied. In order to validate these predictions empirically within a standard behavioural framework, we present evidence from two laboratory experiments: one which varies the composition of the choice set, and another which varies the size of the choice set. We then demonstrate that the observed behaviour is accurately captured by the NRUM incorporating normalization and compare its performance to existing econometric techniques.

2.1 Proportional Scaling in Choice Stochasticity

The behavioural predictions from normalization stem from the interaction between scale and stochasticity that is present in all RUMs. In a standard RUM without normalization ($\omega = 0$), the scale of the model is set by the ratio of the parameter in the numerator κ to the variance of the stochastic term $\boldsymbol{\eta}$. For instance, when $f(\boldsymbol{\eta})$ is the independent standard normal distribution, the ratio κ/σ determines the variance of a basic Probit model. Therefore from a descriptive standpoint, the effect of introducing normalization is to scale the variance of the RUM in a systematic manner according to the number and value of the alternatives in the choice set.

Some intuition for this scaling and its effect on choice behaviour can be observed in Figure 6. Consider two choice sets (**A** and **B**) in which the difference between the valuations ($v_1 - v_2$) in each set are constant, but the alternatives in set **A** are of higher *absolute* value. In a standard RUM (with constant variance), the choice probability of the higher-valued alternative in each set (v_1) would be equivalent since the model is identified only for differences in valuation (i.e. the vectors $\mathbf{v}_\mathbf{A}$ and $\mathbf{v}_\mathbf{B}$ are equi-distant from the 45° line). But under normalization there is now an important distinction between these two choice sets. Due to the geometry of normalization, the higher-valued alternative in each set is scaled more (in absolute terms) than the lower. Therefore the decision-maker will

make more choice errors in the higher-valued set (**A**) because it has a smaller *relative* ratio than the lower-valued set (**B**). In other words, the degree of choice stochasticity between two alternatives is proportional to the magnitude of the valuations, and normalization implements choice behaviour compatible with a formulation of Weber’s Law (Weber, 1834; Glimcher, 2011).

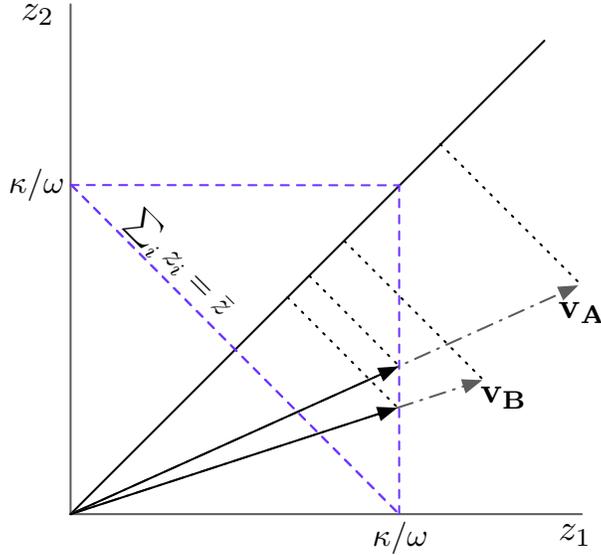


Figure 6: Choice stochasticity scales with value of the choice set.

2.2 Substitution Patterns and Violations of IIA in Trinary Choice

Further predictions arise as the choice set is expanded. Consider the introduction of a third alternative to a choice set, depicted graphically in Figure 7 under the assumption of a linear constraint (i.e. a simplex in \mathbb{R}_+^3). Under normalization, the neural activity for each alternative is re-scaled to abide by the bound constraint, proportionate to the value of the third alternative, v_3 . As before, the highest-valued alternative is scaled more than the second, therefore the absolute magnitude between z_1 and z_2 shrinks when the third alternative is added. This scaling continues as the value of a third alternative is increased, implementing a pattern of substitution in the choice probabilities (Figure 8). As the value of the third alternative is increased, this decreases the odds (the ratio $\frac{P_1}{P_2}$) that alternative 1 is chosen relative to alternative 2.¹⁰

The types of substitution patterns induced by normalization depart from the IIA axiom in a testable manner. To understand how, let us return to

¹⁰This is true if η has a constant variance, but can remain true if this assumption is relaxed.

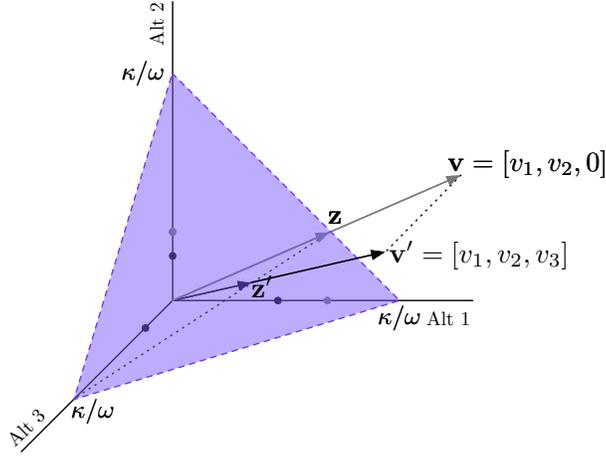


Figure 7: The introduction of a third alternative rescales activity according to the normalization equation. The small circles represent the coordinates of the normalized vector \mathbf{z} .

the NRUM without normalization (i.e. $\omega = 0$ in equation 9). In the standard RUM framework, substitution patterns are determined by the distribution of the error vector $\boldsymbol{\eta}$. For instance, if we assume the η_i are independently distributed extreme value (type 1), then the RUM takes the form of the familiar Multinomial Logit model (McFadden, 1978),

$$P_i = \frac{e^{v_i}}{\sum_n e^{v_n}},$$

and the ratio $\frac{P_i}{P_j} = \frac{e^{v_i}}{e^{v_j}}$ does not depend on the additional alternatives $k \neq i, j$ in the choice set. This property defines the IIA axiom (Luce, 1959).

Though it is now well-documented that IIA is violated in a variety of empirical and theoretical choice environments (Rieskamp, Busemeyer and Mellers, 2006), the axiom still provides a useful benchmark for exploring substitution patterns. In particular, incorporating normalization in the NRUM predicts patterns of violation which depend on the distribution of $\boldsymbol{\eta}$. As we have just seen, an important element of these patterns is the decrease in the probability ratio $\frac{P_1}{P_2}$, and is present even when the error distribution is extreme value.¹¹ This result can be seen clearly via simulation in Figure 9A: the ratio of $\frac{P_1}{P_2}$ declines as the subjective value of the third alternative increases.

If we instead assume a normal distribution for the stochastic elements, the pattern of IIA violation predicted by normalization changes in an important way (Figure 9B). As the value of the third alternative increases, the ratio $\frac{P_1}{P_2}$ still

¹¹Formally, $\frac{\partial \frac{P_i}{P_j}}{\partial v_k} = \frac{\partial e^{\frac{v_i - v_j}{\sum_n v_n}}}{\partial v_k} < 0, \forall k \neq i, j$.

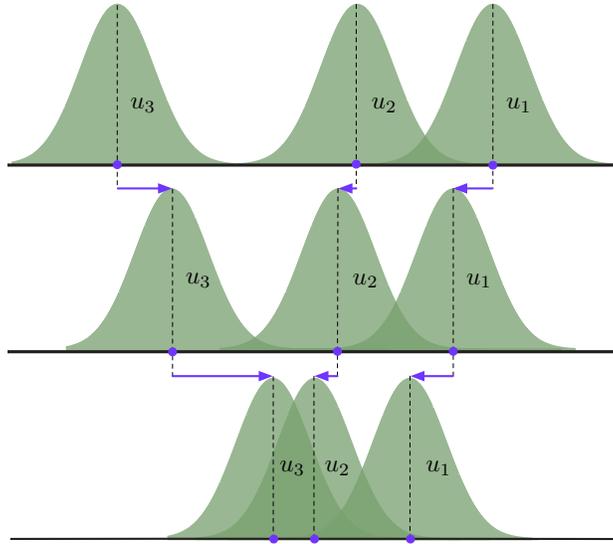


Figure 8: Divisive normalization in an NRUM for trinary choice. An increase in the value of alternative 3 affects the relative likelihood of choosing between alternatives 1 and 2.

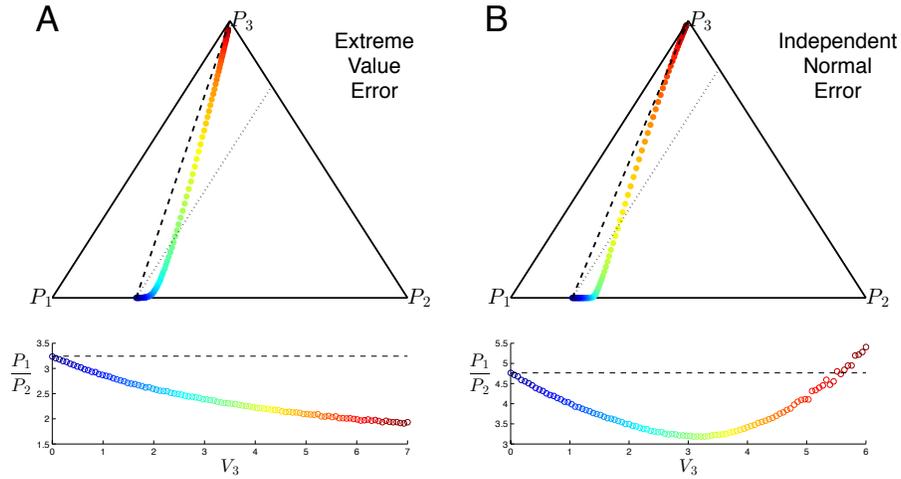
declines, but only until the third alternative starts capturing significant choice probability from the first and second alternatives. Then the ratio $\frac{P_1}{P_2}$ increases and the first alternative is chosen relatively more often (Figure 9B). Intuition for this result can be found in Figure 8. As the density of u_3 shifts to the right, it overlaps substantially with the density of u_2 before that of u_1 . This pulls more probability from the second alternative being chosen relative to the first, sending the probability of choosing the second alternative quickly to zero. This results in a substitution pattern between alternatives which has a characteristic “u-shape.”

2.3 Alternative existing RUM specifications and their relationship to normalization

Adaptations of the RUM framework have been previously proposed in order to relax the IIA axiom, all working through the specification of the error distribution. In an important respect, a distributional specification is also what we are proposing here. If one simply multiplied the denominator of the normalization equation (9) through to the right-hand side of the inequality, the distribution of the error η could be re-defined appropriately. However this distribution would carry with it a precise specification – with precise predictions for behaviour – that differs from current approaches.

One common method for capturing IIA violations is the Multinomial Pro-

With Normalization



Without Normalization

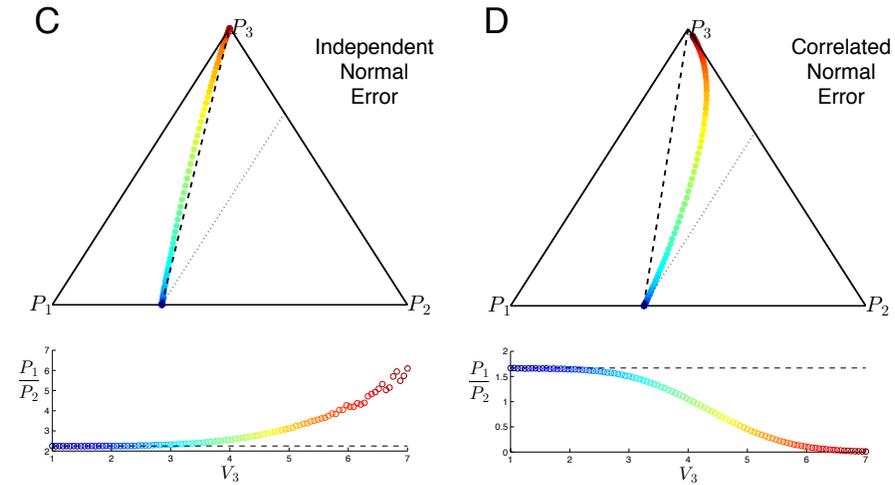


Figure 9: Simulated choice probabilities from an NRUM with (A,B) and without (C,D) normalization. The IIA property is denoted by the dashed line. The iso-probability line for P_2 is denoted by the dotted line. Panel A: Error term is distributed extreme value. Panels B,C: Error term is distributed independent multivariate normal. D: Error term is distributed multivariate normal with correlation.

bit model (MNP; Hausman and Wise, 1978) in which the stochastic portion of utility is distributed multivariate normal with a fully-parameterized covariance matrix. This flexible specification has led to the assertion that the model can “represent any substitution pattern” (Meyer, 1991), though this is clearly too strong of a claim. For instance, substitution patterns which arise from a relationship between the variance of the model and the magnitude of valuations in the choice set (i.e. heteroskedasticity as function of \mathbf{v}) can not be captured by the MNP. Since the lack of a closed-form expression for the MNP prohibits a definitive statement of which substitution patterns the MNP can and cannot capture, the ultimate test of any such model will have to be empirical.¹² In the trinary choice experiment which follows, we will indeed empirically verify that the MNP model performs worse than a model with normalization in the context of individual choice behaviour.

The most general class of model for capturing IIA violations is the Mixed Logit model (McFadden and Train, 2000). This general approach introduces a mixing distribution over the parameters of the extreme value distribution and can approximate any RUM (as well as some models which violate regularity) and the normalization model we are proposing.¹³ However any mixture model still requires the researcher to specify a particular specification for the mixing distribution, with important consequences. As (under)stated by Train (2009), “the researcher’s task is simply to find variables and a mixing distribution that capture the [non-idiosyncratic] parts of utility, namely the parts that are correlated over alternatives or heteroskedastic”.

By incorporating divisive normalization in an NRUM, we accomplish this task directly by imposing a *structural* specification on the relationship between alternatives. We constrain the model by grounding it in the neural constraints underlying decision, and let these constraints induce patterns of substitution between alternatives. We find these patterns are modulated by a combination of at least two factors: the distribution of the stochastic terms (including its covariances), and the presence of normalization. The predictions of an NRUM with and without normalization therefore allow observed substitution patterns to distinguish between these factors in an empirical specification.

There is one additional observation to make about the relationship between the NRUM with normalization and existing RUMs. The regularity condition, which states that the probability of choosing an alternative cannot increase as more alternatives are added to a choice set, is a necessary condition for a random utility representation (Becker, DeGroot and Marschak, 1963). Interestingly, we find that some parameterizations of the NRUM with normalization violate regularity, and we provide a proof in Appendix C for the case of the extreme value distribution. Some intuition for this result can be garnered from the trinary choice simulations in Figures 9A and 9B. Note that for very low valuations of

¹²The lack of a closed-form expression for the multi-variate normal distribution requires that the MNP model be estimated using simulation methods (Train, 2009).

¹³Define $\boldsymbol{\eta} \equiv \boldsymbol{\eta}^* + \lambda \mathbf{e}$ with the elements of \mathbf{e} i.i.d. extreme value. Denote $\mathbf{u}^* = Z(\mathbf{v}) + \boldsymbol{\eta}^*$, such that $\mathbf{u} = \mathbf{u}^* + \lambda \mathbf{e}$. Conditional on $\boldsymbol{\eta}^*$ and \mathbf{z} , the model is logit and the unconditional probability P_i is mixed logit.

the third alternative (where the probability of it being chosen is virtually zero), the probability of choosing the second alternative increases in absolute terms, and quite starkly. We will examine this prediction in more detail in section 3.2 when we present our experiment which examines the effect of increasing the size of the choice set.

3 Choice Experiments

We now turn to the empirical evidence for the substitution patterns predicted by divisive normalization when the value composition of a choice set is varied. To test some of these predictions in a familiar behavioural setting, Louie, Khaw and Glimcher (2013) developed a two-stage valuation and choice task involving common snack-food items. The first stage of the experiment elicited the valuation of each item using an incentive-compatible auction mechanism (Becker, DeGroot and Marschak, 1964), yielding an *independent* measure of value for each of the items in the experiment. The second stage of the experiment had subjects choose a single alternative from choice sets composed from the items. Therefore the composition and/or size of these choice sets in the second stage provides variation that can identify the substitution patterns that arise from the normalization model.

In this article, we will analyze both the original dataset from Louie, Khaw and Glimcher (2013) for trinary choice sets and present data from a new experiment which varies the size of the choice set. In both of these analyses, we will use the behavioural measures of value recorded in the first stage in place of the neural measures proposed in the NRUM of section 2.¹⁴ The variation in these behavioural measures of valuation across choice sets will allow us to examine the substitution patterns that emerge in the choice probabilities, both non-parametrically and with a parametric estimation of the normalization model. This will allow us compare the normalization model to alternative econometric specifications for behavioural datasets.

3.1 Trinary Choice Experiment

In the first stage of the experiment, 40 subjects performed 60 bid trials to establish subject-specific valuations for the array of experimental goods. In each bid trial, subjects viewed a picture of a single snack food item and reported their maximum willingness-to-pay for that item (\$0-\$4 in \$0.01 increments). To implement incentive compatibility, if a bid trial was chosen for realization at the end of the session, the outcome was determined via an auction mechanism (Becker, DeGroot and Marschak, 1964). Each of 30 items was shown twice in a randomized order, and multiple bids for the same item were highly correlated ($\rho = 0.9105$, $p < 0.001$, Figure 10). As a result, we ranked each item according

¹⁴Empirical evidence that subjective value activity correlates with BDM bids can be found in Plassmann, O’Doherty and Rangel (2007).

to its mean bid (for each subject, Figure 10) and used the mean bid in the rest of our analysis.

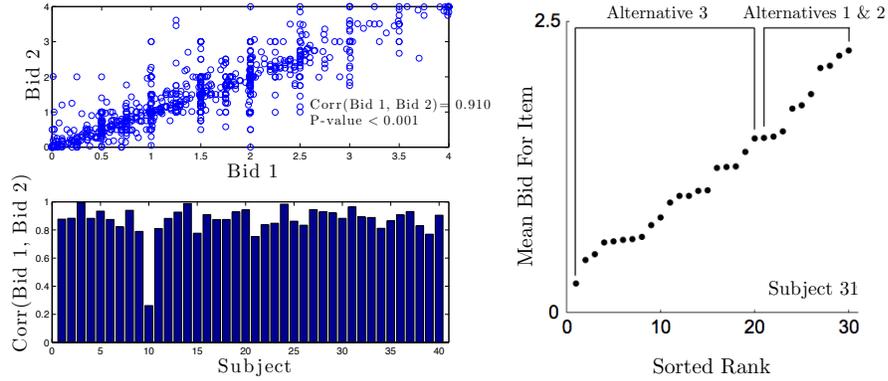


Figure 10: Top Left: The two bids for each snack food alternative from the first stage of the experiment, for all subjects. Bottom Left: The correlation coefficient for each subject. Right: Ranked bids from an example subject. Tertiary choice sets were constructed by drawing two items from the top 10 ranked items (alternatives 1 & 2) and one uniformly from the lower 20 ranked items (alternative 3).

In the second stage of the experiment, subjects performed choice trials from 250 unique choice sets composed of three alternatives. Each trinary set randomly sampled (without replacement) two alternatives from the 10 highest-valued items (termed alternatives ‘1’ & ‘2’ according to their rank order), and one alternative sampled uniformly from the 20 lowest-valued items (termed alternative ‘3’). Each presented triplet was therefore unique. For each alternative in the choice set, we denote the mean of the two bids for the alternative with the variable b_i , where i is the rank-order of the item in a particular set. To facilitate comparison across subjects, we will also report the bid for each item *relative* to the mean bid for all items (within a subject) as the variable \tilde{b}_i .

If a choice trial was chosen for realization, subjects received whichever good they selected in that trial. Subjects were requested to fast for four hours before the experiment and required to stay in the lab one hour after completion of the session, during which time the only food they could consume was any item received from the experiment. Along with a possible snack item, each subject received a \$40 show-up fee and was additionally endowed with \$4 for use in the bid trials. At the conclusion of the experiment, a single trial from the session (bid or choice) was randomly selected for realization.

3.1.1 Trinary Choice Analysis

We begin our analysis of the trinary choice dataset with a comparison of the observed choice frequencies across subjects. Due to heterogeneity in preferences, each subject’s bid distribution for the 30 items varies widely. For some subjects, bids for the items that comprise alternative 3 can be substantially lower than those for alternatives 1 and 2, while for other subjects alternative 3 bids may be very close to those for 1 and 2. This across-subject variation provides our first test of the divisive normalization model. Under normalization, we should expect to see lower odds of choosing alternative 1 vs. 2 for subjects with higher valuations of alternative 3.

This is indeed the case. Figure 11 reports the observed sample probabilities ($\hat{P}_1, \hat{P}_2, \hat{P}_3$) for each subject, ordered by the average \tilde{b}_3 . Not only does the choice ratio between alternatives 1 and 2 decrease, but the probability that alternative 2 is chosen (\hat{P}_2) increases by ~ 0.19 as the average \tilde{b}_3 increases ($p < 0.043$, Table 1).¹⁵

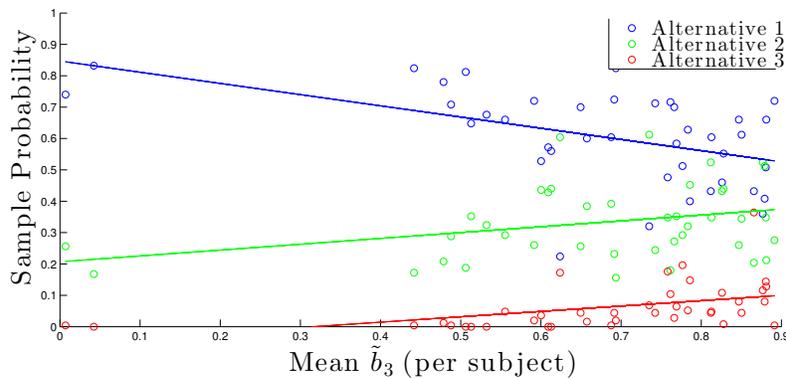


Figure 11: The sampled choice probabilities of alternatives 1 (blue), 2 (green), & 3 (red), ordered by the mean \tilde{b}_3 for each subject. The results for each subject align vertically. A linear fit for each alternative is also depicted.

While an across-subject result is suggestive, the crux of the normalization result occurs within each choice trial depending on the particulars of the choice set and each subject’s preferences. For instance, we note that the above pattern could also arise because subjects with larger average \tilde{b}_3 tend to have smaller value differences between the top two alternatives ($\tilde{b}_1 - \tilde{b}_2$). An analysis introduced in Louie, Khaw and Glimcher (2013) controlled for such issues at the level of the choice trial by coding the bid of alternative 3 relative to the mean value of the presented targets on each trial ($b_3 - \frac{b_1 + b_2}{2}$). This metric alleviates issues with differing bid ranges across subjects by focusing solely on the composition of the set on each trial. The analysis dropped the small number of

¹⁵This result is robust to the two outlier subjects with small average \tilde{b}_3 .

Table 1: Results from an OLS regression of the sample choice probabilities on the average \tilde{b}_3 , for each subject.

	Constant	Mean \tilde{b}_3
\hat{P}_1	0.847 (0.073)	-.357 (0.104)
\hat{P}_2	0.207 (0.065)	0.186 (0.092)
\hat{P}_3	-0.054 (0.038)	0.170 (0.054)

trials in which alternative 3 was chosen, and a logit choice model was fit to the remaining data segregated into five bins of $(b_3 - \frac{b_1+b_2}{2})$. Relative preference is therefore inferred from a binary choice between the alternatives 1 and 2 (within a bin of alternative 3), controlling for differences in valuations of the top two alternatives, and the magnitude of the logit parameter (the slope of the logistic function) gives a measure of the ratio $\frac{P_1}{P_2}$ at the mean value for $b_1 - b_2$.¹⁶

Figure 12 reports the results from the above analysis. We observe a significant decrease in the magnitude of the logit parameter as the value of alternative 3 approaches the top two alternatives. There is also evidence for a “u-shape” in relative choice probabilities: the probability of selecting alternative 1 initially decreases by from 0.67 to 0.57 at the mean of $b_1 - b_2$, but increases back to 0.68 at sufficiently large values of alternative 3. This substitution pattern is consistent with a normalization model with a normal distribution for $\boldsymbol{\eta}$ as described in section 2.2. Further within-subject results showing evidence of normalization can be found in appendix A.

3.1.2 Model Estimation and Comparison

We now turn to estimation of the normalization model (equation 9) and a comparison with two alternative specifications: the Multinomial Probit (MNP) and Multinomial Logit (MNL) models. The following specification of the choice probabilities nests each of these models under certain restrictions (see Table 2):

$$P_i(\mathbf{b}) = \int \mathbb{1} \left[\frac{\kappa(b_i - b_j)}{\sigma + \omega(\sum_{n=1}^3 b_n^\beta)^{\frac{1}{\beta}}} > \eta_j - \eta_i, \quad \forall j \neq i \right] f(\boldsymbol{\eta}) d\boldsymbol{\eta}. \quad (10)$$

The key parameter is ω , which measures the influence of the choice set composition on overall variance. If we restrict $\omega = 0$, normalization is absent and the model reduces to familiar forms depending on one’s choice of an error distribution. For instance, equation 10 becomes the Probit model when $f(\cdot)$ is the independent standard normal density and κ is appropriately normalized. When

¹⁶Essentially, this method assumes that IIA holds “locally” within a bin in order to assess whether it is violated “globally” over bins.

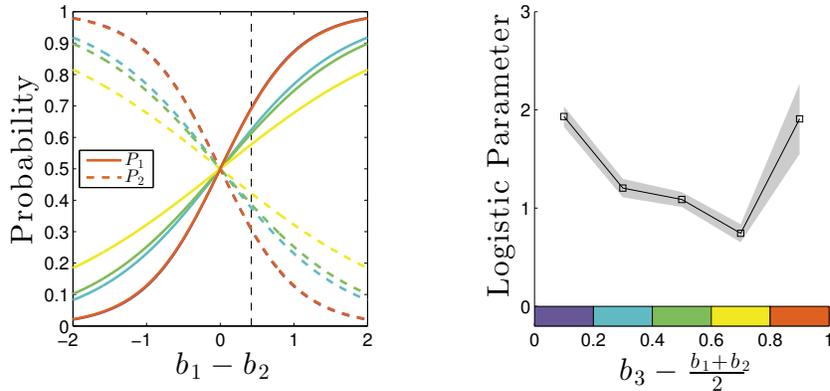


Figure 12: Logistic fit of binary choice between targets, for the range of alternative 3 bids normalized on each trial and segregated into five bins. The relative preference between two alternatives in a choice set is given by the ratio of the probabilities for a given $b_1 - b_2$ (e.g. the mean of $b_1 - b_2$, dashed line). The “u-shaped” relative choice probability can be observed in the estimates of the logistic parameter over the quintiles of alternative 3.

$f()$ is the multivariate normal density, equation 10 yields the MNP model fully parameterized with the constant variance term σ and the cholesky factorization of the covariance matrix, differenced with respect to the first alternative, $(l_{2,1}^1 \ l_{2,2}^0)$. If $f()$ is Extreme Value density, we have the Multinomial Logit.¹⁷

However when ω is unrestricted, normalization is present and the variance in the model is a function of the value of alternatives in the choice set. We estimate three different specifications with this property. First, to verify that substitution patterns in the dataset can be captured solely by normalization (in conjunction with the normal distribution), we restrict $\beta = 1$ and the covariance matrix to be independent. Second, to determine how the normalization model, by itself, compares to the MNP we relax our restriction on β . In this specification, σ represents the constant portion of the variance, while the term governed by ω and β represents the normalized portion of the variance, for a total of three parameters. Finally, we estimate a full specification (with an unrestricted covariance matrix) that nests both the normalization model and the MNP.¹⁸ In each of these specifications, if we find our estimate of ω to be positive we can

¹⁷We implement the Logit via the `clogit` command in Stata (for estimating $\hat{\kappa}$) with its customary normalization of the variance of η_i to $\frac{\pi^2}{6}$, $\sigma = 1$, and $\omega = 0$. This yields the estimate $\hat{\sigma} = \frac{\pi^2}{6} \frac{1}{\hat{\kappa}}$.

¹⁸In the estimation, the parameters σ , ω , and β were restricted to be positive. For large values of β , the approximation $\lim_{\beta \rightarrow \infty} (\sum_{n=1}^3 x_n^\beta)^{\frac{1}{\beta}} = \max\{x_n\}$, was used for numerical purposes. When the covariance matrix is independent, we integrate the multivariate normal density using a Gauss-Hermite approximation. To estimate the covariance matrix parameters, we simulated the multivariate normal density with the GHK algorithm.

reject a model with constant variance in favor of a model with normalization.

Each specification is estimated on the full pooled sample (250 trials \times 40 subjects = 10000 observations) via Maximum Likelihood and results are reported in Table 2.¹⁹ In each of the normalization specifications, we find our estimate of ω to differ significantly from zero ($p < 0.001$) and the log-likelihood of our normalization specifications decreases substantially compared to the nested Probit specification. Our estimate of β is less than 1, but not significantly different from 1 at the 1% level ($p = 0.040$). These results suggest that the composition of the choice set affects the degree of variance in choice behaviour, and therefore the relative probabilities, in a manner consistent with the normalization model presented in section 2.²⁰

Table 2: Estimates of the normalization model and alternative specifications for the trinary choice experiment. Standard errors are in parentheses. Restricted parameters are denoted in red.

	MN Logit	Probit	MN Probit	Normalization		
κ	1	1	1	1	1	1
$\hat{\sigma}$	0.880 (0.036)	0.996 (0.004)	1.493 (0.053)	0.114 (0.024)	0.012 (0.022)	0.121 (0.049)
$\hat{\omega}$	0	0	0	0.177 (0.007)	0.412 (0.014)	0.183 (0.051)
$\hat{\beta}$	1	1	1	1	513.7	0.723 (0.135)
$\hat{l}_{1,1}$	-	1	1	1	1	1
$\hat{l}_{2,2}$	-	0.866	0.792 (0.033)	0.866	0.866	0.780 (0.044)
$\hat{l}_{2,1}$	-	0.5	0.431 (0.022)	0.5	0.5	0.454 (0.029)
LL	-8190.54	-8153.55	-7969.81	-7944.39	-7881.25	-7803.86

To determine whether the MNP can capture the substitution patterns we ob-

¹⁹We also compute estimates using the relative bids \bar{b} instead of the absolute bids b_i and report them in appendix B. The pattern of results is similar, with the lone exception being an improvement in the performance of the MNP model relative to the restricted (three-parameter) normalization model.

²⁰In the normalized specifications, there is a ridge in the log-likelihood function for large values of β due to the approximation used for the norm. Some intuition can be found in Figure 18 in Section 4. When β is large, there is little change in the geometric shape implied by normalization as β varies, therefore little change in the choice probabilities for a typical choice set in our experiment. The large estimate of β (in one of our specifications) should therefore not be interpreted as a point estimate. To calculate standard errors of $\hat{\omega}$ and $\hat{\sigma}$, we hold $\hat{\beta}$ fixed at its estimated value. To give a measure of the variance of $\hat{\beta}$, we calculate that $\beta = 476.63$ would yield a likelihood ratio statistic where $\chi^2(LR, 1) = 0.05$.

serve in this dataset, we also estimated the MNP model ($\omega = 0, \beta = 1$) with an unrestricted covariance matrix, for a total of three parameters ($\sigma, l_{2,1}$, and $l_{2,2}$). The log-likelihood from this model was lower than each of our restricted normalization specifications, and a non-nested likelihood ratio test (Vuong, 1989) rejects the MNP in favour of the three-parameter normalization specification with an independent covariance matrix ($p < 0.001$). To gauge where the improvement in fit is coming from, we also report the fitted probabilities for the normalized model and the MNP in Figure 13 over the quintiles of \tilde{b}_3 . *While the MNP is able to capture a decreasing probability ratio, to do so it misses badly on the overall level of variance in the data.* This is not the case for the normalized model which matches all of the probabilities well. When both the normalization parameters and the covariance matrix are unrestricted, we see a further improvement in fit. Clearly, the choice behaviour we are observing is more accurately captured by a model which includes normalization.

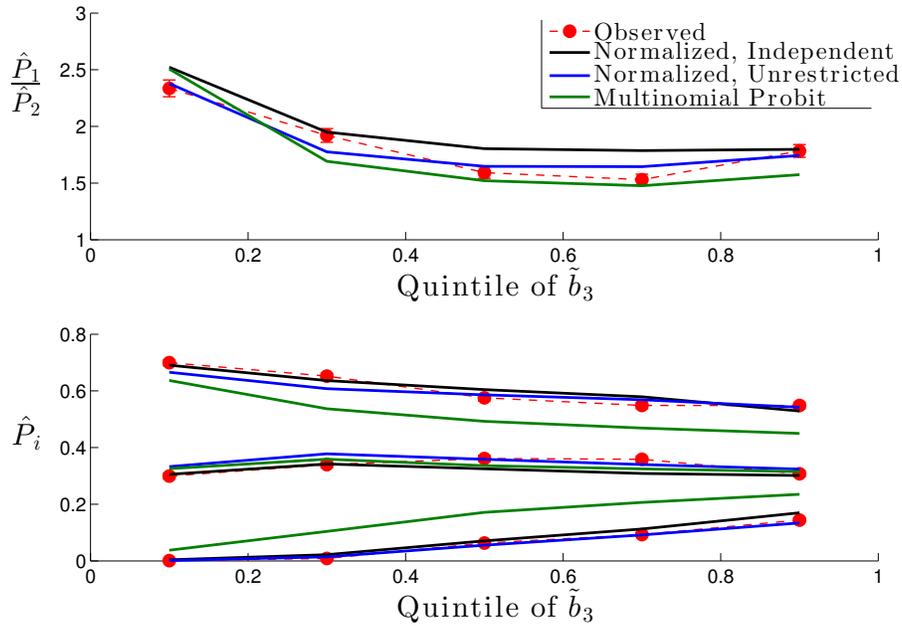


Figure 13: The probabilities and odds of alternative 1 vs. 2 computed for the quintiles of (relative) bids for alternative 3. Also depicted are the fitted probabilities from estimation of the normalized model (both with an independent and unrestricted covariance matrix) and the MNP model with an unrestricted covariance matrix.

3.2 Set Size Experiment

The behavioural predictions of divisive normalization naturally extend to larger choice sets. In a typical RUM like the Probit model, increasing the size of the choice set will decrease the probability of choosing the existing alternatives. This is because the new alternative is chosen with positive probability and the probability of each alternative must sum to 1. However in a normalized RUM, there is a second effect to consider.

Increasing the number of choice alternatives changes the total value of the choice set and requires neural activity to be re-proportioned (Figure 7 and Figure 14). As in trinary choice, the relative choice probabilities shift and choice becomes more stochastic; in particular the likelihood of choosing the highest valued alternative is decreased. Therefore as alternatives are added to a choice set, we should observe *more* stochasticity than is predicted in a model with a constant variance, and observe a particular pattern in the probabilities of choosing the remaining alternatives. In sections 3.2.1 and 3.2.2, we present a new behavioural experiment designed to test these predictions.

As mentioned in section 2.3, normalization also predicts violations of regularity under some parameterizations of the model and for some sets of alternatives. The difference between the decision variables for the highest and second highest-ranked alternative (u_1 and u_2) can decrease quite quickly when multiple low-value alternatives are added to the choice set. With a small variance for the stochastic elements of η , and low-enough value alternatives added to the set, this can generate an absolute increase in the probability of choosing the second-ranked alternative (Appendix C). In section 3.2.3 we will test for this prediction in the data.

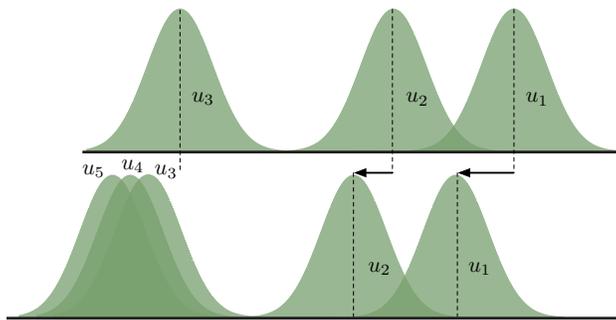


Figure 14: Increasing the size of the choice set decreases the relative probability of choosing the highest ranked alternative.

3.2.1 The Set Size Task

To try and capture the behavioural effects of increasing set size, we developed a two-stage experimental setup based on the trinary choice experiment. In the

initial valuation stage, 30 subjects performed 60 bid trials where they reported their maximum willingness-to-pay for 30 individual snack food items. For each subject, items were then ranked according to their mean bid value, and the 10 highest-valued and 10 lowest-valued items were selected for inclusion in the second stage of the experiment.

In the second stage, subjects performed 270 choice trials from choice sets of varying sizes. On each trial t , a choice set consisted of two high-value alternatives (again, denoted alternatives ‘1’ & ‘2’ in rank-order) and a variable number of randomly chosen low-valued items for a total set size of $N_t \in \{2, 4, 6, 8, 10, 12\}$ items. Choice sets were constructed to include varying target value differences, with 45 different trials in each set size condition.

After completion of the session, subjects were required to stay in the lab for one hour, during which time the only food they could consume was any item received from the experiment. A single experimental trial was chosen for realization. If a bid trial was selected, the outcome was determined via BDM auction; if a choice trial was selected, subjects received whichever item was chosen in that trial. Payment consisted of a \$40 showup fee, a \$4 endowment for use in bid trial realization, and a single possible snack food item. In the analysis that follows, the bids for each subject were normalized by the subject’s mean bid over all items to facilitate pooling the data.

3.2.2 Set Size Analysis

As one might expect, increasing the size of the choice set decreases the probability that the highest ranked alternative is chosen. In our largest set size of 12 alternatives, the decrease in probability is as high as 20% (Figure 15). In order to determine if this decrease was larger than predicted by a RUM with constant variance (Probit), we estimated our normalized NRUM from equation (9) with the only difference lying in the number of choice elements, N_t , now varying over trials. This yields the choice probabilities on trial t ,

$$P_{i,t}(\tilde{\mathbf{b}}) = \int \mathbb{1} \left[\frac{\kappa(\tilde{b}_{i,t} - \tilde{b}_{j,t})}{\sigma + \omega(\sum_{n \in N_t} \tilde{b}_{n,t}^\beta)^{\frac{1}{\beta}}} > \eta_{j,t} - \eta_{i,t}, \quad \forall j \neq i \right] f(\boldsymbol{\eta}_t) d\boldsymbol{\eta}_t. \quad (11)$$

As in the trinary case, the distribution $f(\cdot)$ was assumed independent normal, κ was appropriately normalized, and both the parameters σ and ω were restricted to be positive.²¹ The resulting maximum likelihood estimates for the entire pooled sample are reported in Table 3 for three nested specifications.²²

Again, we observe that in the unrestricted specification, the estimate of ω is significantly different from zero ($p < 0.001$). This suggests that *we are observing choice behaviour in which the size of the choice set affects the degree of variance in the model*, and therefore the relative probabilities. In order to quantify the

²¹A specification which unrestricted the covariance matrix is infeasible for this dataset since the number of alternatives, therefore the size of the covariance matrix, changes on each trial.

²²The results for the normalization model do not change if we use b_i instead of \tilde{b}_i , and the Probit model fits slightly worse with $\hat{\sigma} = 1.21$ and a log-likelihood of -10066.2 .

Table 3: Maximum likelihood estimates for set size experiment. Restricted parameters are denoted in red.

	Probit	Normalization	
$\hat{\sigma}$	0.924 (0.025)	0.924 (0.002)	0.001 (0.040)
$\hat{\omega}$	0	0.001 (0.010)	0.442 (0.022)
$\hat{\beta}$	1	1	18.85 (1.961)
LL	-9969.76	-9969.76	-9723.77

effect of set size, we have reported the fitted probabilities from our estimates in Figure 15 evaluated at the average bid values of the choice sets. The specification with normalization captures the relative choice probabilities in the data, with the slight exception of under-predicting the probability of choosing alternative 3 in large set sizes.²³ This is in contrast to a RUM without normalization (i.e. Probit). In order to match the choice probabilities for all alternatives with a constant variance over all sets, it over-predicts the probability of choosing alternative 1 and under-predicts the probability of choosing alternative 2 & 3. From these fitted probabilities, we can observe that for a set size of 12 alternatives, the highest ranked alternative is chosen 10% less than predicted by a RUM with a constant variance.

Table 4: Probability of choosing highest-ranked alternative (evaluated at average bids in each choice set)

	$P_{N=2}$	$P_{N=12}$	$P_{N=12} - P_{N=2}$
Data	0.67	0.47	0.20
Probit ($\hat{\sigma} = 0.924$)	0.72	0.57	0.15
Normalization ($\hat{\sigma} = 0.001, \hat{\omega} = 0.442, \hat{\beta} = 18.85$)	0.69	0.49	0.20

²³The increase in the predicted probability of choosing alternative 3 arises from the construction of our choice set. Alternatives 3-12 were randomly sampled from a set of 10 items, which means that smaller choice sets were more likely to have a larger disparity between alternative 3 and the remaining alternatives. Notice in Figure 15 that the difference between average bid values for each alternative is initially large, but shrinks as set size increases.

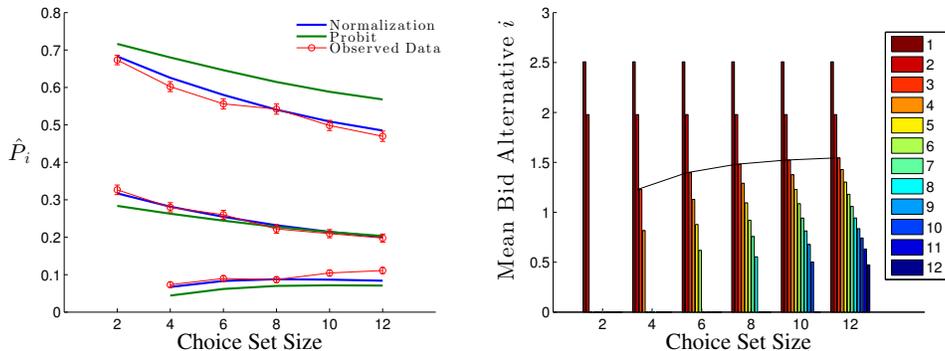


Figure 15: (Left) The sample probabilities that alternative 1, 2, & 3 are chosen (from top to bottom) over the range of choice set sizes, and the fitted probabilities derived from model estimates. (Right) The probabilities are evaluated at the average bids in each choice set.

3.2.3 Regularity Violations

As noted in section 2.3, the NRUM with normalization predicts violations of regularity for a small range of parameter values (see also Appendix C). Specifically, the probability of choosing the second ranked alternative should increase when the number of alternatives added to a choice set are all of low value and the variance in the model is small.

To investigate this prediction, we report the probability of choosing the second ranked alternative for four subsets of the data (in blue) in Figure 16. The subsets are generated by selecting the trials (from each subject) in which all of the low-valued alternatives were in the 50th, 55th, 60th, and 70th percentile of bids for that subject. Ideally, we would examine subsets lower than this range since this is where we would expect the largest number of regularity violations, but the design of our experiment severely underweights sampling from this range of items. This significantly reduces the number of trials which fall into these subsets and limits the power of our dataset to identify the type of regularity violations predicted by normalization. As can be observed in Figure 16, there are some cases in which the average probability of choosing the second alternative does increase with the size of the choice set, but these results are not statistically significant.

Whether or not a model with normalization obeys regularity is a matter of model classification. If a model with normalization violates regularity, it can not be formally rationalized with a random utility representation (Becker, DeGroot and Marschak, 1963). From an empirical standpoint this should not be surprising since violations of regularity have been previously observed in the literature (Rieskamp, Busemeyer and Mellers, 2006). In any event, an NRUM with normalization can capture (at least some) violations of regularity in a model that maintains the mechanism of the random utility model, if not its

formal representation.

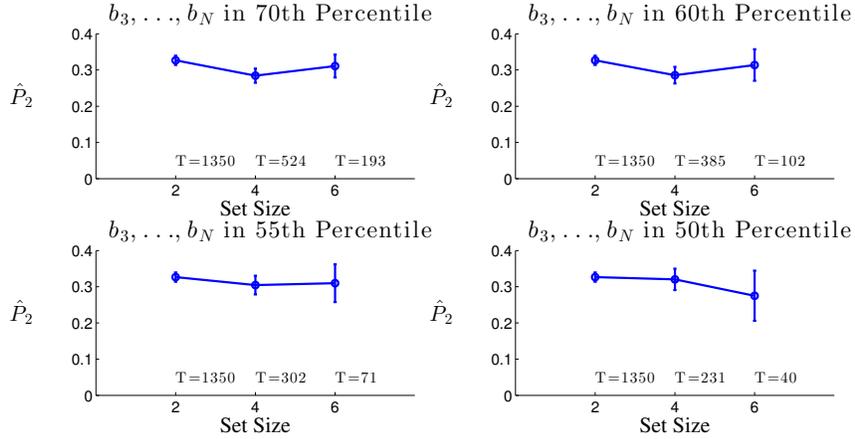


Figure 16: The probability of choosing the second ranked alternative for four subsets of the data (in blue) for choice sets of increasing size. Each subset is composed of all trials for each subject in which the distracters were less than the 50th, 55th, 60th, and 70th percentile of the distracter bids for that subject. The number of trials (over all subjects) in the subset are reported, and the probabilities calculated from the remaining trials are depicted (in black).

4 Normative Role of Normalization

In the two experiments just presented, we observed choice behaviour that can be strictly deemed inconsistent or perhaps even welfare-decreasing. The ability of our subjects to select their highest-ranked alternative was degraded when both the composition and the size of the choice set was altered, and a choice model which incorporates normalization captures this and other behavioural patterns better than alternative econometric approaches. This observation raises the natural question of why a neural system – selected over tens of millions of years of evolution – would exhibit such adverse choice performance.

Recall that the neurobiological constraint on decision-making that we identified in section 1 has two implications. The first is that neural activity is stochastic, both empirically and theoretically, and this stochasticity can be shown to be optimal given a capacity constraint (Woodford, 2014). The second is a bound on neural activity in decision-making regions of the brain. The existence of these constraints re-focuses the normative question slightly. How should the brain encode the valuation of choice alternatives given a neurobiological constraint on the precision and resolution with which it can encode quantities, and why might this encoding scheme follow the form of divisive normalization?

In section 1.2.1, we noted that the normalization computation has a normative basis in sensory systems because it reduces the redundant information contained in natural sensory stimuli (Wainwright, Schwartz and Simoncelli, 2001). The computation achieves this increase in efficiency because it encodes each input relative to a set of inputs. This significantly reduces the information (therefore resources) required to encode, for example, the luminance of a light in the bounded activity of a finite set of neurons, and is consistent with the normative criterion in systems neuroscience, the *efficient coding hypothesis* (Barlow, 1961).

However, this constrained encoding scheme (like all constrained optimal solutions) has consequences. While the encoding itself is efficient given the constraint, it can still lead to errors in perception, and ultimately, choice. Consider the image known as the “Checkerboard Illusion” depicted in Figure 17. If required to decide which of the two regions “A” and “B” appears brighter, one could not be faulted for choosing region “B”. In fact, the luminance of the two regions is precisely the same. Region “B” is encoded relative to nearby locations that fall in a shadow, and is therefore normalized to a lesser extent than region “A”. Therefore “B” appears brighter even though it is of the same (objective) luminance, potentially resulting in a choice error.

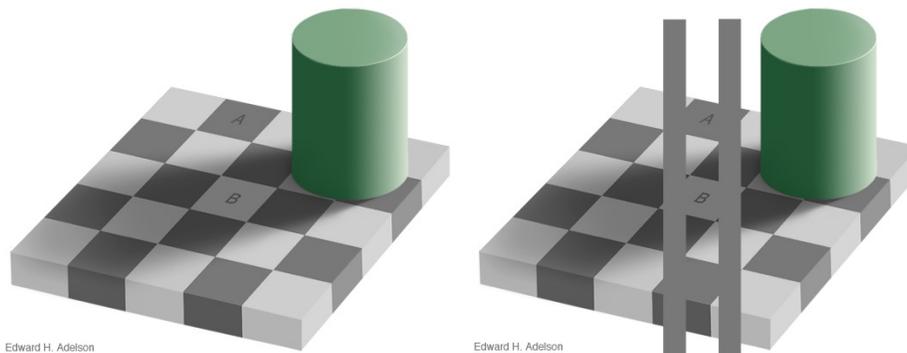


Figure 17: Left: The “Checkerboard Illusion.” Reproduced with permission from <http://persci.mit.edu/gallery> Right: fixed luminance bars reveal the luminance of regions “A” and “B” to be equal.

In the context of economic choice, our normative observation essentially amounts to an argument that divisive normalization leads to as few choice errors as possible given the above neurobiological constraints.²⁴ In the simple binary choice depicted in Figure 18.A, the probability of making a choice error

²⁴Our focus on choice errors follows the economic literature which has previously used errors as a metric of optimality (Robson, 2001; Rayo and Becker, 2007; Robson and Samuelson, 2010). While we do not argue explicitly in terms of the evolutionary criteria of “expected loss” (Netzer, 2009), which incorporates the magnitude of a choice error, in the appendix we show that the general divisive normalization computation does allow a degree of monotonicity in the transformation of subjective values, such that “more valuable” choice sets lead to fewer errors than less valuable choice sets. Of course, a full normative treatment of the neural

is determined by the distance from the vector \mathbf{v} to the 45° line. Formally, let $\mathbf{M}_\iota \mathbf{v} = \mathbf{v} - \frac{\iota \cdot \mathbf{v}}{\iota \cdot \iota} \iota$ denote the image of the projection of the vector \mathbf{v} to the unit vector ι . In the absence of normalization, the length $\|\mathbf{M}_\iota \mathbf{v}\|$, relative to the variance of the stochastic term $\boldsymbol{\eta}$, would determine the probability of choosing a lower-valued alternative.

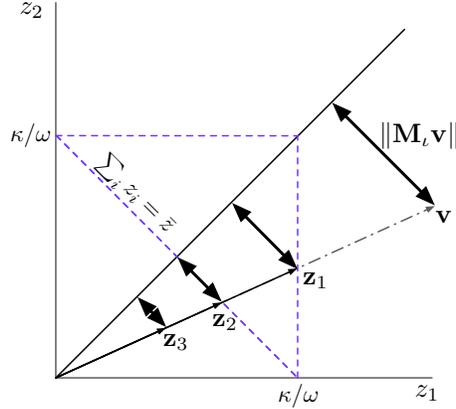


Figure 18: Divisive normalization scales \mathbf{v} to the bound constraint. This reduces the probability of making an error as much as possible, given the constraint.

The normative role of divisive normalization lies in the fact that it scales \mathbf{v} directly to the bound constraint (\mathbf{z}_2), not to some point interior (\mathbf{z}_3). Therefore the length $\|\mathbf{M}_\iota \mathbf{z}\|$ is made as large as possible given the bound on neural activity, and the probability of making an error is reduced as much as possible. It is then clear that the magnitude of β also has normative implications. When $\beta > 1$, normalization maps the vector \mathbf{v} to a family of curves that lies outside the linear constraint. Therefore $\|\mathbf{M}_\iota \mathbf{z}\|$ grows, with the smallest possible probability of error when $\beta = \infty$, and \mathbf{v} is mapped to the square (\mathbf{z}_1).

As we have described in section 2, the constraint implemented by normalization can lead to seemingly sub-optimal choice behaviour. For instance, it can generate intransitivities arising from stochasticity and a decrease in the probability of choosing the highest-valued alternative. As more choice alternatives are added, or the value of some choice alternatives are modulated, this adverse effect on choice behaviour is amplified yielding the substitution patterns we observed empirically in section 3.

However we should emphasize that these patterns predicted in choice behaviour might be viewed as optimal given the neurobiological constraints facing the decision-making system. Since normalization scales valuations to the boundary of the constraint (along a ray to the origin), it preserves relative valuations

decision-making problem would require a complete accounting of the costs of decision-related errors to the chooser and the metabolic costs of relaxing the constraints we identify — on an evolutionary timescale.

and limits choice errors as much as possible. Viewed this way, the adverse choice behaviour that results is not due to an inefficient or mal-adapted neural architecture, but is in fact an efficient implementation in response to a neurobiological constraint on the capacity of the choice representation.

5 Conclusion

In the study of the human brain and its functional relationship to behaviour, it has long been recognized that the biology of neural systems is subject to constraints imposed by finite energy resources and is shaped by the statistical properties of the environment we live in. As a result, the organization and form of neural computation have evolved such that information about the objective world is approximated and/or compressed into quantities that can be encoded in finite, stochastic, neural activity. It is only natural to examine what these constraints imply for choice behaviour, and whether theories of economic behaviour need incorporate these insights. In this article, we pursue this goal directly.

It has been established that measurements of neural quantities in the human brain (which we term subjective value) correlate with all tested incentive-compatible methods for measuring value and predicting choice behaviour (Bartra, McGuire and Kable, 2013; Levy and Glimcher, 2011; Clithero and Rangel, 2013). This has led to the development of neuroeconomic models which take neural quantities as their primitives, and allow for neurobiological constraints to be imposed on a model of economic behaviour (Webb et al., 2013; Fehr and Rangel, 2011). In this article, we extend one such model, the Neural Random Utility Model (Webb et al., 2013), to incorporate a canonical neural computation known as *divisive normalization*. This computation serves to implement a relative coding of information and has been observed across all sensory modalities, a vast range of species, and recently in decision-making areas of the primate brain (Carandini and Heeger, 2012; Louie, Grattan and Glimcher, 2011).

When incorporated in a neuroeconomic model, divisive normalization induces a degree of context-dependency in choice behaviour. We demonstrate that divisive normalization serves to suppress, or normalize, the subjective value of a particular choice alternative relative to the composition and size of the choice set. This creates a relationship between the neural quantities used to implement a decision, yielding predictions about the relative choice probabilities for a given choice set. The effects of this relationship include increased stochasticity, particular patterns of violation of the Independence of Irrelevant Alternatives axiom, and a decreased likelihood of choosing a higher-valued alternative. We present evidence for these patterns in choice behaviour from two laboratory experiments, one which varies the composition of the choice set and one which varies its size, and demonstrate that a model which incorporates normalization captures observed behaviour more accurately than existing econometric methods.

Finally, we address the normative role of the relative valuations induced by

the divisive normalization computation. In sensory systems, normalization has been demonstrated to reduce the resources required to encode the statistical properties of our natural environment (Schwartz and Simoncelli, 2001; Wainwright, Schwartz and Simoncelli, 2001), however in the domain of decision-making its normative role has remained unclear (Louie, Khaw and Glimcher, 2013). That normalization predicts the adverse choice behaviour noted above – together with the fact that we observe this behaviour in our choice experiments – brings this question to the forefront. Why would the brain implement a computation that can lead to adverse choice performance? Our answer is that the existence of a neurobiological resource constraint requires some degree of normalization of subjective value for the purposes of comparing alternatives in a choice set. Neural activity is subject to a bound constraint on its representational capacity, and we demonstrate that a neuroeconomic choice model employing normalization scales valuations to this constraint, preserves relative valuations, and provides an efficient means for the brain to represent valuations because it yields the fewest possible choice errors.

All together, our results speak to the role of incorporating insights from biology into economic discourse. While developing an understanding of the brain is intriguing in its own right, the goal of neuroeconomic research should be predicting novel patterns in choice behaviour and producing a normative explanation for the behaviour we observe (Bernheim, 2009; Glimcher, 2011). We would argue that divisive normalization, incorporated in a Neural Random Utility Model, does just this. Our results emphasize both the positive advances offered by choice models grounded in neuroscience, and the normative role neurobiological constraints can play in the study of decision-making.

References

- Barlow, H B.** 1961. “Possible principles underlying the transformation of sensory messages.” In *Sensory Communication*. MIT Press.
- Bartra, Oscar, Joseph T McGuire, and Joseph W Kable.** 2013. “The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value.” *NeuroImage*, 76: 412–427.
- Becker, Gordon M, Morris H DeGroot, and Jacob Marschak.** 1963. “Stochastic models of choice behavior.” *Behavioral Science*, 8(1): 41–55.
- Becker, Gordon M, Morris H DeGroot, and Jacob Marschak.** 1964. “Measuring utility by a single-response sequential method.” *Behavioral Science*, 9(3): 226–232.
- Ben-Akiva, Moshe E, and Steven R Lerman.** 1985. *Discrete Choice Analysis. Theory and Application to Travel Demand*, MIT Press.

- Bernheim, B Douglas.** 2009. “On the potential of neuroeconomics: a critical (but hopeful) appraisal.” *American Economic Journal: Microeconomics*, 1(2): 1–41.
- Block, H D, and Jacob Marschak.** 1960. “Random Orderings and Stochastic Theories of Responses.” *Cowles Foundation Discussion Papers*.
- Bogacz, Rafal, Eric-Jan Wagenmakers, Birte U Forstmann, and Sander Nieuwenhuis.** 2010. “The neural basis of the speed-accuracy trade-off.” *Trends in neurosciences*, 33(1): 10–16.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2012. “Saliency theory of choice under risk.” *The Quarterly Journal of Economics*, 127(3): 1243–1285.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer.** 2013. “Saliency and Consumer Choice.” *Journal of Political Economy*, 121(5): 803–843.
- Brownstone, David, and Kenneth Train.** 1999. “Forecasting new product penetration with flexible substitution patterns.” *Journal of Econometrics*, 89: 109–129.
- Busemeyer, Jerome R, and James T Townsend.** 1993. “Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment.” *Psychological Review*, 100(3): 432–459.
- Caplin, Andrew, and Mark Dean.** 2014. “Revealed Preference, Rational Inattention, and Costly Information Acquisition.” *Working Paper*.
- Carandini, Matteo, and David J Heeger.** 2012. “Normalization as a canonical neural computation.” *Nature Reviews Neuroscience*, 13(1): 51–62.
- Clithero, John A, and Antonio Rangel.** 2013. “Informatic parcellation of the network involved in the computation of subjective value.” *Social cognitive and affective neuroscience*, 1–14.
- Debreu, Gerard.** 1960. “Individual Choice Behavior: A Theoretical Analysis by R. Duncan Luce.” *The American Economic Review*, 50(1): 186–188.
- DeShazo, J R, and German Fermo.** 2002. “Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency.” *Journal of Environmental Economics and Management*, 44(1): 123–143.
- Dorris, Michael C, and Paul W Glimcher.** 2004. “Activity in posterior parietal cortex is correlated with the relative subjective desirability of action.” *Neuron*, 44(2): 365–378.
- Falmagne, J C.** 1978. “A representation theorem for finite random scale systems.” *Journal of Mathematical Psychology*, 18(1): 52–72.

- Fehr, Ernst, and Antonio Rangel.** 2011. “Neuroeconomic Foundations of Economic Choice—Recent Advances.” *Journal of Economic Perspectives*, 25(4): 3–30.
- Glimcher, Paul W.** 2005. “Indeterminacy in Brain and Behavior.” *Annual Review of Psychology*, 56(1): 25–56.
- Glimcher, Paul W.** 2011. *Foundations of Neuroeconomic Analysis*. Oxford University Press.
- Gold, Joshua I, and Michael N Shadlen.** 2007. “The neural basis of decision making.” *Annual Review of Neuroscience*, 30: 535–574.
- Hartline, H K, and F Ratliff.** 1957. “Inhibitory interaction of receptor units in the eye of Limulus.” *The Journal of general physiology*, 40(3): 357–376.
- Hausman, JA, and David 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.
- Heeger, David J.** 1992. “Normalization of cell responses in cat striate cortex.” *Visual Neuroscience*, 9(2): 181–197.
- Hess, Stephane, and John M Rose.** 2009. “Allowing for intra-responder variations in coefficients estimated on repeated choice data.” *Transportation Research Part B: Methodological*, 43: 708–719.
- Hodgkin, A L, and A F Huxley.** 1952. “A quantitative description of membrane current and its application to conduction and excitation in nerve.” *The Journal of Physiology*, 117(4): 500–544.
- Huber, Joel, John W Payne, and Christopher Puto.** 1982. “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis.” *Journal of Consumer Research*, 9(1): 90–98.
- Iyengar, S S, and M R Lepper.** 2000. “When choice is demotivating: can one desire too much of a good thing?” *Journal of Personality and Social Psychology*, 79(6): 995–1006.
- Kable, Joseph W, and Paul W Glimcher.** 2007. “The neural correlates of subjective value during intertemporal choice.” *Nature Neuroscience*, 10(12): 1625–1633.
- Kable, Joseph W, and Paul W Glimcher.** 2010. “An ‘As Soon As Possible’ Effect in Human Intertemporal Decision Making: Behavioral Evidence and Neural Mechanisms.” *Journal of Neurophysiology*, 103: 2513–2531.
- Karni, E, and M L Vierø.** 2013. “‘Reverse Bayesianism’: A choice-based theory of growing awareness.” *The American Economic Review*, 103(7): 2790–2810.

- Kiani, Roozbeh, Timothy D Hanks, and Michael 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.
- Levy, Dino, and Paul W Glimcher.** 2011. “Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain.” *The Journal of Neuroscience*, 31(41): 14693–14707.
- Levy, Dino J, and Paul W Glimcher.** 2012. “The root of all value: a neural common currency for choice.” *Current Opinion in Neurobiology*, 22(6): 1027–1038.
- Levy, Ifat, J Snell, A J Nelson, Aldo Rustichini, and Paul W Glimcher.** 2010. “Neural Representation of Subjective Value Under Risk and Ambiguity.” *Journal of Neurophysiology*, 103(2): 1036–1047.
- Levy, Ifat, Stephanie C Lazzaro, Robb B Rutledge, and Paul W Glimcher.** 2011. “Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing.” *The Journal of Neuroscience*, 31(1): 118–125.
- Louie, Kenway, and Paul W Glimcher.** 2010. “Separating value from choice: delay discounting activity in the lateral intraparietal area.” *The Journal of Neuroscience*, 30(16): 5498–5507.
- Louie, Kenway, Lauren E Grattan, and Paul W Glimcher.** 2011. “Reward value-based gain control: divisive normalization in parietal cortex.” *The Journal of Neuroscience*, 31(29): 10627–10639.
- Louie, Kenway, Mel W Khaw, and Paul W Glimcher.** 2013. “Normalization is a general neural mechanism for context-dependent decision making.” *Proceedings of the National Academy of Sciences*, 110(15): 6139–6144.
- Luce, R. Duncan.** 1959. *Individual Choice Behaviour*. John Wiley and Sons, Inc.
- Lyu, Siwei.** 2011. “Dependency reduction with divisive normalization: justification and effectiveness.” *Neural computation*, 23(11): 2942–2973.
- McFadden, D.** 1978. “Modelling the choice of residential location.” *Cowles Foundation Discussion Papers*.
- McFadden, Daniel.** 1974. “Conditional Logit Analysis of Qualitative Choice Behavior.” In *Economic Theory and Mathematical Economics*. Academic Press, Inc.
- McFadden, Daniel.** 2001. “Economic choices.” *The American Economic Review*, 91(3): 351–378.

- McFadden, Daniel, and K Train.** 2000. "Mixed MNL models for discrete response." *Journal of Applied Econometrics*, 15: 447–470.
- McFadden, Daniel L.** 2005. "Revealed stochastic preference: a synthesis." *Economic Theory*, 26(2): 245–264.
- Meyer, Robert J.** 1991. "Probabilistic Models of Consume Choice Behavior." In *Handbook of Consumer Behavior*. Prentice Hall, Inc.
- Milosavljevic, M, J Malmaud, A Huth, Christof Koch, and Antonio 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.
- Netzer, Nick.** 2009. "Evolution of Time Preferences and Attitudes toward Risk." *The American Economic Review*, 99(3): 937–955.
- Padoa-Schioppa, Camillo.** 2011. "Neurobiology of Economic Choice: A Good-Based Model." *Annual Review of Neuroscience*, 34(1): 333–359.
- Padoa-Schioppa, Camillo, and J A Assad.** 2006. "Neurons in the orbitofrontal cortex encode economic value." *Nature*, 441(7090): 223–226.
- Pastor-Bernier, Alexandre, and Paul Cisek.** 2011. "Neural correlates of biased competition in premotor cortex." *The Journal of Neuroscience*, 31(19): 7083–7088.
- Plassmann, Hilke, John P O’Doherty, and Antonio Rangel.** 2007. "Orbitofrontal cortex encodes willingness to pay in everyday economic transactions." *The Journal of Neuroscience*, 27(37): 9984–9988.
- Platt, Michael L, and Paul W Glimcher.** 1999. "Neural correlates of decision variables in parietal cortex." *Nature*, 400(6741): 233–238.
- Qamar, Ahmad T, R James Cotton, Ryan G George, Jeffrey M Beck, Eugenia Prezhdo, Allison Laudano, Andreas S Tolias, and Wei Ji Ma.** 2013. "Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization." *Proceedings of the National Academy of Sciences*, 110(50): 20332–20337.
- Ratcliff, Roger.** 1978. "A theory of memory retrieval." *Psychological Review*, 85(2): 59–108.
- Rayo, Luis, and Gary S Becker.** 2007. "Evolutionary Efficiency and Happiness." *Journal of Political Economy*, 115(2): 302–337.
- Revelt, David, and Kenneth Train.** 1998. "Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level." *Review of Economics and Statistics*, 80(4): 647–657.

- Ribrault, C, K Sekimoto, and A Triller.** 2011. “From the stochasticity of molecular processes to the variability of synaptic transmission.” *Nature Reviews Neuroscience*, 12(7): 375–387.
- Rieskamp, J., Jerome R Busemeyer, and B.A. Mellers.** 2006. “Extending the bounds of rationality: evidence and theories of preferential choice.” *Journal of Economic Literature*, 44(3): 631–661.
- Robson, Arthur J.** 2001. “The biological basis of economic behavior.” *Journal of Economic Literature*, 39(1): 11–33.
- Robson, Arthur J, and L Samuelson.** 2010. “The Evolutionary Foundations of Preferences.” In *Handbook of Social Economics*. 221–310. North Holland.
- Roitman, JD, and Michael N Shadlen.** 2002. “Response of Neurons in the Lateral Intraparietal Area during a Combined Visual Discrimination Reaction Time Task.” *The Journal of Neuroscience*.
- Rorie, Alan E, Juan Gao, James L McClelland, and William T Newsome.** 2010. “Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey.” *PloS one*, 5(2): e9308.
- Schwartz, O, and E P Simoncelli.** 2001. “Natural signal statistics and sensory gain control.” *Nature Neuroscience*, 4(8): 819–825.
- Shadlen, Michael N, and Roozbeh Kiani.** 2013. “Decision making as a window on cognition.” *Neuron*, 80(3): 791–806.
- Shadlen, Michael N, and William T Newsome.** 1998. “The variable discharge of cortical neurons: implications for connectivity, computation, and information coding.” *The Journal of Neuroscience*, 18(10): 3870–3896.
- Shannon, C E.** 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal*, 27(3): 379–423.
- Simon, H A.** 1979. “Rational Decision Making in Business Organizations.” *The American Economic Review*, 69(4): 493–513.
- Sims, Christopher A.** 2003. “Implications of rational inattention.” *Journal of Monetary Economics*, 50(3): 665–690.
- Sinz, Fabian, and Matthias Bethge.** 2013. “Temporal Adaptation Enhances Efficient Contrast Gain Control on Natural Images.” *PLoS Computational Biology*, 9(1): e1002889.
- Smith, Alec, B Douglas Bernheim, Colin F Camerer, and Antonio Rangel.** 2013. “Neural Activity Reveals Preferences Without Choices.” *NBER Working Paper*.

- Soltani, Alireza, Benedetto De Martino, and Colin F Camerer.** 2012. “A range-normalization model of context-dependent choice: a new model and evidence.” *PLoS Computational Biology*, 8(7): e1002607.
- Sugrue, Leo P, Greg S Corrado, and William T Newsome.** 2004. “Matching Behavior and the Representation of Value in the Parietal Cortex.” *Science*, 304: 1782–1787.
- Thaler, Richard.** 1985. “Mental accounting and consumer choice.” *Marketing science*, 4(3): 199–214.
- Thaler, Richard H.** 1999. “Mental Accounting Matters.” *Journal of Behavioral Decision Making*, 12: 123–206.
- Train, Kenneth E.** 2009. *Discrete Choice Methods with Simulation*. . 2nd ed., Cambridge University Press.
- Train, Kenneth E, and Clifford Winston.** 2007. “Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers.” *International Economic Review*, 48(4): 1469–1496.
- Tusche, Anita, Stefan Bode, and John-Dylan Haynes.** 2010. “Neural Responses to Unattended Products Predict Later Consumer Choices.” *The Journal of Neuroscience*, 30(23): 8024–8031.
- Tversky, Amos, and Itamar Simonson.** 1993. “Context-Dependent Preferences.” *Management Science*, 39(10): 1179–1189.
- Vuong, Quang H.** 1989. “Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses.” *Econometrica: Journal of the Econometric Society*, 57(2): 307.
- Wainwright, Martin J, O Schwartz, and E P Simoncelli.** 2001. “Natural Image Statistics and Divisive Normalization: Modeling Nonlinearities and Adaptation in Cortical Neurons.” In *Statistical Theories of the Brain*. MIT Press.
- Webb, R, Paul W Glimcher, Ifat Levy, S Lazzaro, and Robb B Rutledge.** 2013. “Neural Random Utility and Measured Value.” *SSRN Working Paper*.
- Webb, Ryan.** 2013. “Dynamic constraints on the distribution of stochastic choice: Drift Diffusion implies Random Utility.” *SSRN Working Paper*.
- Weber, E H.** 1834. *E. H. Weber: On the Tactile Senses (with translation of De Tactu)*. Experimental Psychology Society, New York.
- Woodford, Michael.** 2012. “Prospect Theory as Efficient Perceptual Distortion.” *The American Economic Review*, 102(3): 41–46.

Woodford, Michael. 2014. “Stochastic Choice: An Optimizing Neuroeconomic Model.” *American Economic Review, Papers and Proceedings*, 104(5): 495–500.

APPENDIX FOR ONLINE PUBLICATION

A Further Within-Subject Analysis of Trinary Choice Experiment

The small number of observations for each subject ($T = 250$) limits the statistical power of a within-subject analysis, but support for the normalization model can still be found. For instance, we should expect to see a smaller ratio of alternative 1 chosen relative to alternative 2 when the value of alternative 3 is large. After a median split of the relative bid \tilde{b}_3 for each subject, the ratio $\frac{\hat{P}_1}{\hat{P}_2}$ is smaller for the upper median of b_3 for 34 of 40 subjects, with 13 of them significantly smaller at the 0.05 level.

The normalization model can also be estimated at the level of each subject to gauge heterogeneity in context-dependent choice behaviour. Normalization yields a prediction about *how much* context-dependent choice behaviour we should observe for a given range of choice sets. For instance, if trinary choice sets were composed of a Porsche, a Ferrari, and some lowly snackfood, we should not expect to see much substitution as the snack food alternative is varied because it is such a relatively lower-valued alternative. Therefore the heterogeneity in the preferences of subjects for the various snack food items in our sample can be used as an additional test of the model.

The resulting maximum likelihood estimates for each subject are in Figure 19.A, in which the estimate $\hat{\omega}$ is multiplied by the average value of the choice set to give a sense of scale relative to $\hat{\sigma}$. The statistical significance of the estimate $\hat{\omega}$ is calculated via a likelihood ratio test from the restricted model where $\omega = 0$.

For each subject, the estimated model either portions variance to σ or to the normalization term governed by ω . This division of our subject pool can also be observed if we repeat the logistic analysis from Louie, Khaw and Glimcher (2013) on just those subjects for which $\hat{\omega}$ is significant, and those for which it is not, reported in Figure 19.B. The subset of subjects which yield a significant $\hat{\omega}$ exhibit a strong contextual effect on the choice probabilities, in contrast to the remaining subjects (who still exhibit the characteristic increase induced by the multivariate normal distribution).

So what to make of the observation that some subjects appear to be exhibiting contextual effects while others are not? Recall that the behavioural predictions from the divisive normalization equation arise because the difference in value between alternatives $b_1 - b_2$ is scaled down by the valuation of the third alternative, b_3 , in the denominator. If there is little exogenous variation of b_3 in our choice sets, or if the magnitude of $b_1 - b_2$ is small to begin with, then we should not observe much effect on the relative choice probabilities. For instance, in Figure 19.C we highlight the ordered bids of three subjects, two of

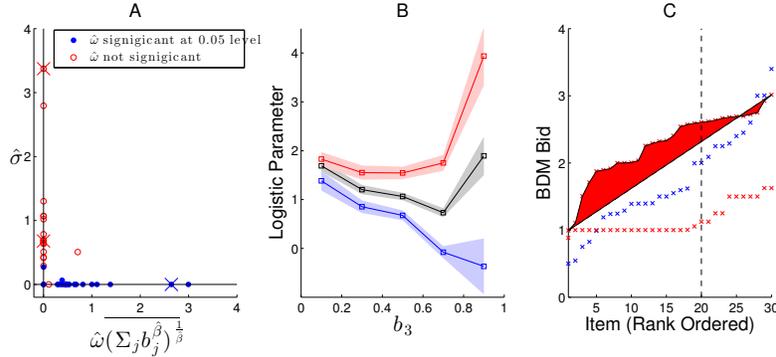


Figure 19: (A) The estimates of σ vs. ω for each subject. Heterogeneity in the sample is clearly present, with some subjects demonstrating more normalization in their choice probabilities (in blue) than others (in red). (B) The logistic fit re-estimated for each sub-sample of subjects. (C) Heterogeneity is predicted by a normalization model depending on a subject’s range of valuations over the content of the item set (shown for the 3 subjects denoted with an ‘x’). The red polyhedron depicts the departure of the observed bids (for one subject) from a line between the minimum and maximum bid.

which (in red) the model identifies as not normalizing ($\omega \approx 0$). One of these subjects has a small range of valuations of alternative 3 (b_3), while the other has a small range of valuations of alternative 1 & 2 (b_1, b_2), relative to the normalizing subject (in blue). Note that the bid spectrum of the normalizing subject has roughly equal range for both distractors and targets, therefore is roughly linear compared to the non-normalizing subjects.

To test this possible explanation for the discrepancy in our subject sample, we constructed a measurement of how much variation is present in valuations of the item set. For each item, we calculated the square of the difference between the rank-ordered valuations and a line which ranges from the minimum bid to the maximum bid. A subject for which the target and distractor set shared a similar range would exhibit a small measurement of variation and a large amount of normalization. A subject for which either the low-value items or high-value items yielded differing ranges of valuations would exhibit a large measurement and little normalization. We then averaged this metric over the two subsets of subjects defined by our regression results, and found it was significantly larger ($p < 0.05$) for those subjects which did not exhibit normalization. This suggests that the lack of normalization behaviour observed in a subset of our subjects resulted from a set of choice alternatives which were not conducive to generating choice set effects (in those subjects).

B Alternative model estimation and comparison using the relative bids (\tilde{b})

In Figure 5 we report the results from the estimation of equation (10) using the relative bids (\tilde{b}_i) in place of the absolute bids (b_i) for the trinary choice set experiment. Results are similar to the those reported in section 3.1 with the exception of the ordering of the log-likelihoods for the MNP model and the three parameter normalization model.

Table 5: Estimates of the Normalization model and alternative specifications for the trinary choice experiment. Standard errors are in parentheses. Restricted parameters are denoted in red.

	MN Logit	Probit	MN Probit	Normalization		
κ	1	1	1	1	1	1
$\hat{\sigma}$	0.746 (0.041)	0.872 (0.015)	1.383 (0.050)	0.018 (0.039)	0.013 (0.045)	0.579 (0.094)
$\hat{\omega}$	0	0	0	0.204 (0.011)	0.411 (0.025)	0.274 (0.037)
$\hat{\beta}$	1	1	1	1	511.3	33.72 (62.95)
$\hat{l}_{1,1}$	-	1	1	1	1	1
$\hat{l}_{2,2}$	-	0.866	0.843 (0.033)	0.866	0.866	0.732 (0.053)
$\hat{l}_{2,1}$	-	0.5	0.365 (0.022)	0.5	0.5	0.504 (0.033)
LL	-8090.27	-8090.04	-7830.19	-7961.42	-7881.37	-7798.30

C Divisive Normalization and Regularity

We prove that the NRUM with divisive normalization will technically violate the regularity condition for some, but not all, values and parameterizations of the model. Our proof focuses on the extreme value distribution since it has closed-form probabilities. We have not yet characterized the entire range of distributions and range of valuations for which regularity fails, but extensive simulation suggests it is small.

We begin with a definition of Random Utility. Let $P_{i,N}$ denote the probability that a subject chooses alternative i from a finite set N .

Definition 1. (*Random Utilities*) *There exists a random vector $U=(U_1, \dots, U_n)$, unique up to an increasing monotone transformation such that for any $i \in M \subseteq$*

N ,

$$Pr\{U_i \geq U_j, \forall j \in M\} = P_{i,M}.$$

A necessary condition for a random utility representation states that the probability of choosing an alternative decreases as more alternatives are added to the choice set (Block and Marschak, 1960), often referred to as a *regularity condition*.

Definition 2. (*Regularity*) If $L \subseteq M \subseteq N$,

$$P_{i,M} \leq P_{i,L}.$$

We now show that a model with normalization violates the regularity condition for some parameter values, therefore cannot be rationalized by a Random Utility model. We consider two choice sets of different sizes, the three alternative choice set $\{1,2,3\}$, with corresponding valuations $v_1 \geq v_2 \geq v_3 \geq 0$, and the two alternative subset $\{1,2\}$. Since we are assuming an extreme value distribution, the choice probabilities of the normalized model are given by the closed form equation,

$$P_{i,M} = \frac{e^{\kappa \frac{v_i}{\sum_{n \in M} v_n}}}{\sum_{m \in M} e^{\kappa \frac{v_m}{\sum_{n \in M} v_n}}}.$$

We will focus on the probability of choosing the second (ranked) alternative from the set. For our simple choice sets, these probabilities can be re-written

$$P_{2,\{1,2,3\}} = \frac{1}{e^{\frac{\kappa(v_1-v_2)}{v_1+v_2+v_3}} + e^{\frac{\kappa(v_3-v_2)}{v_1+v_2+v_3}} + 1},$$

and

$$P_{2,\{1,2\}} = \frac{1}{e^{\frac{\kappa(v_1-v_2)}{v_1+v_2}} + 1}.$$

We will show regularity is violated as the choice set is expanded from two to three alternatives.

Result 1. *An NRUM which includes divisive normalization is not Regular for some parameter values, therefore cannot be represent by Random Utilities. Specifically, there exists some $\kappa > 0$ such that $P_{2,\{1,2,3\}} > P_{2,\{1,2\}}$.*

Proof. We proceed by contradiction. Suppose $P_{2,\{1,2\}} > P_{2,\{1,2,3\}}$ for all κ . Therefore,

$$\begin{aligned} e^{\frac{\kappa(v_1-v_2)}{v_1+v_2+v_3}} + e^{\frac{\kappa(v_3-v_2)}{v_1+v_2+v_3}} + 1 &> e^{\frac{\kappa(v_1-v_2)}{v_1+v_2}} + 1 \\ e^{\frac{\kappa(v_1-v_2)}{v_1+v_2+v_3}} - \frac{\kappa(v_1-v_2)}{v_1+v_2} + e^{\frac{\kappa(v_3-v_2)}{v_1+v_2+v_3}} - \frac{\kappa(v_1-v_2)}{v_1+v_2} &> 1. \end{aligned}$$

Now we must find values for κ , v_1 , v_2 , and v_3 for which this statement is false. Choose $\kappa = v_1 + v_2$, and $v_3 = v_2 = \frac{1}{2}v_1$. The above statement reduces to

$$\begin{aligned} e^{\frac{(v_1+v_2)(v_1-v_2)}{v_1+v_2+v_3}-(v_1-v_2)} + e^{\frac{(v_1+v_2)(v_3-v_2)}{v_1+v_2+v_3}-(v_1-v_2)} &> 1 \\ e^{\frac{(v_1+v_2)(v_1-v_2)}{v_1+2v_2}-(v_1-v_2)} + e^{-(v_1-v_2)} &> 1 \\ e^{-\frac{1}{8}v_1} + e^{-\frac{1}{2}v_1} &> 1 \end{aligned}$$

As v_1 grows, both elements on the left-hand side tend to zero, and we have a contradiction. \square

D Relaxing σ

Relaxing the assumption that $\sigma = 0$ re-introduces monotonicity into the transformed values (Figure 20). For a given $\sigma > 0$, choice sets with lower total valuations are mapped to a line which intercepts lower than κ/ω .

$$\sum_i z_i = \sum_i \frac{\kappa v_i}{\sigma + \omega \sum_n v_n} = \kappa \frac{\sum_i v_i}{\sigma + \omega \sum_n v_n} < \frac{\kappa}{\omega}. \quad (12)$$

With a positive σ , the length of the vector \mathbf{z} now depends monotonically on the length of the vector \mathbf{v} . This yields the prediction that choice errors are reduced for more valuable choice sets than for less valuable chose sets. However, choice errors increase relative to the case in which $\sigma = 0$, therefore it seems that a neural system for decision-making in *static choice environments* should have a σ near 0 (at least in the case of a constant variance for $\boldsymbol{\eta}$). There is reason to believe this might not be the case in dynamic choice environments in which an average estimate of the choice set, or reference point, can prove useful for achieving optimal choice, suggesting that σ may in fact be a dynamic variable which varies with expectations.

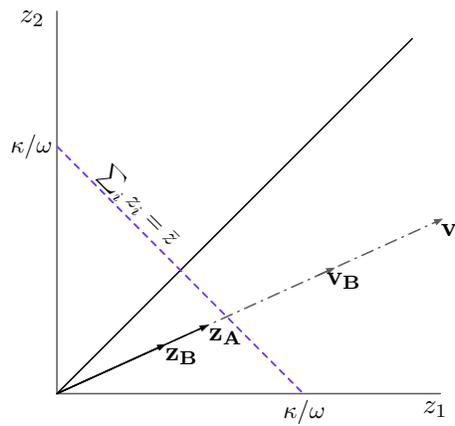


Figure 20: The scaling implemented by normalization when $\sigma > 0$.