Rational Imprecision:
Information-Processing, Neural, and Choice-Rule Perspectives*

Kai Steverson†
Adam Brandenburger‡
Paul Glimcher§

February 26, 2017

Abstract

People make mistakes. A rationally imprecise decision maker optimally balances the cost of reducing mistakes against the value of choosing correctly. We provide three models of rationally imprecise behavior: (1) an information-processing formulation where the costs of reducing mistakes are modeled as the corresponding reduction in Shannon entropy; (2) a neural implementation in terms of a stochastic and context-dependent utility function consistent with how the brain is thought to represent value; and (3) an axiomatic choice-rule characterization. Our main result proves an equivalence between these three models which shows that they are different perspectives on the same behavior. The three perspectives answer, respectively, the questions of why rationally imprecise behavior should arise, how it can be implemented within the brain, and what such behavior looks like. JEL Codes: D87, D81.

*Financial support from NIH grant R01DA038063, NYU Stern School of Business, NYU Shanghai, and J.P. Valles is gratefully acknowledged. We have greatly benefited from discussions with Andrew Caplin, Faruk Gul, Kenway Louie, Paula Miret, Anthony Marley, Wolfgang Pesendorfer, Doron Ravid, Shellwyn Weston, Jan Zimmerman, members of the Glimcher Lab at NYU, and seminar audiences at Columbia, HKUST, NYU Shanghai, PHBS Shenzhen, Princeton, Yale, and Zurich University.

†Corresponding Author. Institute for the Interdisciplinary Study of Decision Making, New York University. Address: 4 Washington Place, Room 809 New York, NY 10003. Email: ksteverson@nyu.edu, Phone: 808-371-6951.

‡Stern School of Business, Tandon School of Engineering, Institute for the Interdisciplinary Study of Decision Making, New York University, New York, NY 10012, U.S.A., adam.brandenburger@stern.nyu.edu, http://www.adambrandenburger.com

§Center for Neural Science, Institute for the Interdisciplinary Study of Decision Making, New York University, New York, NY 10012, U.S.A., paul.glimcher@nyu.edu, http://www.neuroeconomics.nyu.edu/people/paul-glimcher/

1 Introduction

People make mistakes. Even as seemingly simple a task as packing for a trip can lead to mistakes, as when important items are left behind. The chances of forgetting something can be reduced by checking what is in the luggage, but this takes time and effort. This effort may be smaller or larger depending on what is being packed. Choosing the correct articles of clothing can be hard if it is being selected from a drawer of very similar items. Packing a laptop is easy if it is the only computer around. The cost of such time and effort is, presumably, to be offset by the benefits of packing well. These benefits will vary. A long international trip may warrant careful checking and re-checking of a suitcase; an overnight trip may not. In short, decision makers behave optimally when balancing the benefits of choices they make against the costs of reducing the chance of error. We will say that such decision makers are **rationally imprecise**.

In this paper we provide three different models of a rationally imprecise decision maker: an **information-processing formulation**, a **neural implementation**, and a **choice-rule characterization**. The information-processing formulation obtains rationally imprecise behavior by balancing the values of potential outcomes against the cost of reducing stochasticity in choice, formulated entropically. The neural implementation derives rational imprecision from a neural computation, called normalization, which has been found to be involved in choice-related value representation in the brain. The choice-rule characterization yields rational imprecision from axioms which include a relaxation of the classical independence-from-irrelevant-alternatives condition.

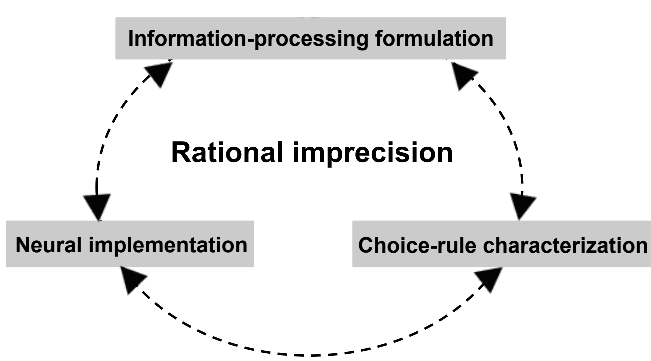

Figure 1

These three models — information-processing, neural, and choice-rule based — represent very different perspectives on rationally imprecise behavior. The main result in this paper is an **equivalence theorem** which establishes that, despite their different vantage points, the three models we build are fully equivalent in terms of the behavior they imply. If observed behavior fits one of the models, then it must fit **all three models**. These models can therefore be seen as offering complementary insights into the phenomenon of imperfectly precise human behavior. See Figure 1.

We next offer more detail on each of the three perspectives. Our **information-processing formulation** views mistakes as resulting from value-reducing stochasticity in choice.[1] Decreasing stochasticity in choices decreases information entropy, which, by the von Neumann-Landauer Principle (Landauer, 1961), involves an inescapable thermodynamic cost. Therefore, a rationally imprecise decision maker faces a trade-off between the the costliness of more mistakes and the — fundamentally energetic — cost of fewer mistakes. In our information-processing model of choice, we specify the cost of decreasing stochasticity to be proportional to the associated decrease in **Shannon entropy** (Shannon, 1948), where the proportionality factor allows for the possibility that the cost may vary depending on features of the choice context. In making an optimal choice, this cost must, of course, be weighed against the expected value of the choice in question.

Our **neural implementation** of rational imprecision proceeds from the idea that mistakes can arise through stochasticity in how the brain represents values. We capitalize on the discovery that activity in certain regions of the brain acts as a representation of value across a wide range of choice tasks (see Levy and Glimcher [2012], for a survey). In work with primates, Louie, Grattan, and Glimcher (2011) showed that this representation of value is well described by a computation called (**divisive**) **normalization** (Carandini and Heeger,

[1]Stochastic decision making is routinely observed in behavioral experiments (e.g., Hey and Orme 1994). At the neuronal level in the brain, stochasticity in response to stimuli is well documented (Shadlen and Newsome, 1994; Churchland et al., 2011).

2012) which is known to describe neural activity in a variety of brain regions and modalities. More recently, normalization has also been demonstrated to explain a variety of observed choice behaviors.[2] (See Appendix A for additional background on normalization.) The normalization computation involves re-scaling the values of choice alternatives by dividing each by a factor that depends on the overall choice set faced. A random error term is then added to the re-scaled values and the alternative with the highest resulting sum is chosen. The error term accounts for observed neurobiological stochasticity in the representation of value and allows for the possibility of mistakes. A larger normalization factor decreases the size of the values relative to the error term and hence increases the chance of a mistake. The normalization factor therefore captures the aspects of the choice context which would make errors more or less prevalent.

Our **choice-rule characterization** gives a direct behavioral description of a rationally imprecise decision maker. Our starting point is the independence-from-irrelevant-alternatives choice rule (IIA) that characterizes the Luce model of stochastic choice (Luce, 1959). IIA says that the ratio of the probabilities of choosing one alternative versus a second is the same in every choice set in which both alternatives appear. IIA allows for mistakes, in that the highest-value alternative is not always chosen, but it does not allow for the context, i.e., the particular choice set, to matter. We introduce choice-set specific functions and relax IIA to require equality of probability ratios across choice sets only after each ratio has been transformed by its associated function. This modification, together with another property carried over directly from the classical IIA rule, defines a choice rule which we call **relaxed IIA**. We go on to provide two closed-form axioms that characterize our rule, so that behavioral testing of the rule is possible.

Returning to our equivalence result, we can think of its three components as answering, respectively, the "why," the "how," and the "what" of our notion of rational imprecision.

[2]Louie, Khaw, and Glimcher (2013) conduct a choice experiment that confirms several predictions of the normalization model. Webb, Glimcher, and Louie (2014) show that normalization explains behavior better than alternative models such as Multinomial Probit, that allow for IIA violations. Glimcher and Tymula (2016) document how normalization can produce many of the behaviors associated with prospect theory.

The information-processing formulation explains why mistakes arise, namely, because of the inescapable physical cost of reducing stochasticity in choice. The neural implementation explains how these mistakes can arise from a mechanism understood to be operative in the brain. The choice-rule characterization says exactly what behavior would be observed.

Our rational imprecision model permits a number of behaviors which are consistent with experimental evidence and which are not well captured by many of the standard existing models. The model can accommodate violations of **regularity,**[3] which occur in well-documented behavior such as the **attraction effect** (Huber, Payne, and Puto, 1982; Simonson, 1989). This ability to accommodate regularity violations distinguishes our model from other well-known models of stochastic choice, such as the random utility model (Thurstone, 1927; Block and Marschak, 1960), the attribute rule (Gul, Natenzon, and Pesendorfer, 2014), and additive perturbed utility (Fudenberg, Iijima, and Strzalecki, 2015). The empirical evidence on two additional behavioral axioms, called **strong** and **weak stochastic transitivity**, was surveyed by Rieskamp, Busemeyer, and Mellers (2006), who found that violations of the first are frequent and occur in a wide variety of circumstances, while violations of the second are rare and occur only in special circumstances. Our model satisfies weak but not strong stochastic transitivity. Also at the empirical level, we show that our model generates a ranking of choice sets in terms of how they affect the stochasticity of choice. This property points to a potential new area for empirical work.

Our equivalence result will work by showing that the three approaches lead to a common formula for the choice probabilities. The mathematics for showing this in the first case, i.e., for our information-processing formulation, comes from basic thermodynamics, subject to reinterpreting some variables. Formally, our information-processing problem is the same as minimizing (Helmholtz) free energy and our choice probability formula is the same as the Boltzmann distribution (see, e.g., Mandl [1988] for the relevant physics). The equivalence is obtained by reinterpreting energy as negative value and (absolute) temperature as our

[3] Regularity violations obtained via the normalization functional form are discussed by Webb, Glimcher, and Louie (2014).

proportionality factor. For our second approach, namely, our neural-normalization implementation of stochastic choice, the key mathematical argument was found by McFadden (1978), although again in a different context, namely, that of estimation of choice models. The mathematics for our third approach, which is our choice-rule characterization of stochastic choice, is new and represents the main technical contribution of the paper.

Swait and Marley (2013), working in mathematical psychology, previously noted the equivalence between our first two approaches, although under a very different interpretation. Instead of an information-processing formulation, they think in terms of an exploitation-exploration trade-off, and, instead of a neural implementation, they see a variation of the Multinomial Logit model. Therefore, our discussion of how the brain may implement the optimal solution to the problem of costly reduction of stochasticity is absent from their work.

The use of Shannon entropy in our information-processing formulation is a point of connection with the rational-inattention literature initiated by Sims (1998, 2003). Recently, the rational-inattention notion has been applied to stochastic choice settings with uncertain values for the alternatives (Matějka and McKay, 2014; Caplin and Dean, 2015). This leads to an information-processing task: Determine the optimal cost to incur in learning about these uncertain values. By contrast, the information-processing task we consider is efficient reduction in the intrinsic stochasticity of choosing among alternatives. At the formal level, our set-dependent proportionality factor distinguishes us from the standard rational-inattention framework. If the proportionality factor is constant across sets, our model is equivalent to a special case of rational inattention.

2 Three Perspectives on Rational Imprecision

We now present three models of rationally imprecise behavior: an information-processing formulation, a neural implementation, and a choice-rule characterization. The main result in this section is an equivalence theorem uniting the three models in terms of the behavior they imply, so that each model provides a different perspective on the same rationally imprecise behavior. The information-processing formulation explains why mistakes arise, namely, because of the costliness of reducing stochasticity in choice. The neural implementation explains how these mistakes can arise from a mechanism believed to be operative in the brain. The choice-rule characterization says exactly what behavior can be observed. In this way, our equivalence theorem allows us to answer the "why," the "how," and the "what" of rationally imprecise behavior.

We begin with the formal framework, which we maintain throughout the paper. Let X be a finite set consisting of all the alternatives from which the decision maker may be able to choose. Let $\mathcal{A}$ be the collection of all non-empty subsets of X, to be thought of as the possible choice sets the decision maker may face. The choice behavior of the decision maker is described by a random choice rule ρ that assigns a full-support probability measure to every choice set $A \in \mathcal{A}$. Formally, a **random choice rule** is a function $\rho : X \times \mathcal{A} \to [0, 1]$ such that $\rho(x, A) > 0$ if and only if $x \in A$ and

$$\sum_{x \in A} \rho(x, A) = 1 \text{ for all } A \in \mathcal{A}.$$

The interpretation is that $\rho(x, A)$ is the probability that the decision maker chooses alternative x when faced with choice set A. (As we will see below, the assumption that ρ is full support is, in fact, a consequence of our information-processing argument.) We let $\mathcal{P}$ denote the set of all random choice rules (for given X).

2.1 Information-Processing Formulation

In our information-processing formulation of rational imprecision, the decision maker balances the expected utility of a given choice rule against the costs involved in reducing stochasticity in choices. The value of alternative x is given by $v(x)$ and the expected utility of choice rule ρ on choice set A is

$$\sum_{x \in A} \rho(x, A) v(x).$$

The information-processing costs of a particular rule ρ come from the reduction in Shannon entropy (Shannon, 1948) relative to the fully stochastic case. Shannon entropy measures the degree of stochasticity in behavior, where a higher degree of stochasticity implies higher entropy. For a random choice rule ρ, the associated Shannon entropy $H(\rho, A)$ on set A is given by

$$H(\rho, A) = -\sum_{x \in A} \rho(x, A) \ln \rho(x, A),$$

where ln denotes natural log.[4] The maximum entropy of $\ln |A|$ is achieved when ρ is fully stochastic and assigns the same probability to each alternative in A. For ease of reference, we write $H_{\max}(A)$ for this value. To arrive at the cost of choice, we add a proportionality factor $F(A)$, which may depend on the choice set A, to the entropy reduction involved. Thus, the total cost of random choice rule ρ on choice set A is given by

$$F(A)(H_{\max}(A) - H(\rho, A)).$$

Combining this expression with the expected value of a choice rule yields our definition of optimal behavior with costly information processing.

Definition 1. A random choice rule ρ has an **information-processing formulation** if there exist functions $v : X \to (0, \infty)$ and $F : \mathcal{A} \to (0, \infty)$ such that for all $A \in \mathcal{A}$:

$$\rho \in \arg\max_{\hat{\rho} \in \mathcal{P}} \left\{ \sum_{x \in A} \hat{\rho}(x, A) v(x) - F(A)(H_{\max}(A) - H(\hat{\rho}, A)) \right\}.$$

Notably, even if we relaxed our definition of a random choice rule to allow for the possibility of assigning probability zero to an available alternative, this turns out never to be optimal in the maximization problem we just set up. This is because the derivative of entropy

[4] Log to base 2 appears in the original definition of entropy, but it is often more conveient to use base e.

become infinite as a probability approaches zero, so that the cost of pushing a probability to zero is always too high. The Appendix contains details.

2.2 Neural-Normalization Implementation

Our neural implementation of rational imprecision comes via the normalization model widely thought to explain choice-related value representation in the brain. The normalization model generates a stochastic and set-dependent utility for each alternative, and the highest utility alternative is then chosen. The utility involves a fixed value $v(x)$ for each alternative x and a normalization factor $F(A)$ for each choice set A. The utility has a non-random component equal to $v(x)/F(A)$, which is the fixed value of x re-scaled by the normalization factor for A. The utility also has a random component, which for each alternative x is a random variable ε_x where these random variables are i.i.d. across different alternatives and each ε_x follows a Gumbel distribution with location 0 and scale 1. The total utility of alternative x in set A is then

$$\frac{v(x)}{F(A)} + \varepsilon_x.$$

We define our normalization model as all behavior that can generated using this utility form. (Note that since the ε_y's in this definition come from a continuous distribution, there is a unique maximizer in the $\arg\max$ with probability one, so that the formula for ρ is well defined.)

Definition 2. A random choice rule ρ has a **neural-normalization implementation** if there exist functions $v : X \to (0,\infty)$ and $F : \mathcal{A} \to (0,\infty)$ such that for all A in $\mathcal{A}$ and x in A:

$$\rho(x, A) = \Pr\left(x = \arg\max_{y \in A} \frac{v(y)}{F(A)} + \varepsilon_y\right),$$

where the ε_y's are i.i.d. and Gumbel with location 0 and scale 1.

Our neural model is distinct from random utility. On the one hand, the presence of the factor $F(A)$ allows a set dependence absent in a standard random utility model. On the other hand, random utility allows for more general assumptions on the error term. The Gumbel distribution we assume does arise in a number of settings — in particular, as the asymptotic distribution of the maximum of a sequence of i.i.d. normal random variables. See, e.g., David and Nagaraja (2003). Many previous studies of normalization, including several of the ones cited in the Introduction, take $F(A)$ to be the sum

$$F(A) = \sigma + \sum_{z \in A} v(z),$$

for some positive constant σ. This functional form is clearly included as a special case of our model, but we do not limit ourselves to it.

2.3 Choice-Rule Characterization

Our choice-rule characterization of rational imprecision relaxes the well-known independence-from-irrelevant-alternatives axiom (IIA) which characterizes the classic Luce model (Luce, 1959) of stochastic choice. IIA requires that

$$\frac{\rho(x, A)}{\rho(y, A)} = \frac{\rho(x, B)}{\rho(y, B)},$$

whenever alternatives x and y are present in both of the sets A and B.

Our relaxation of IIA requires only the following ratio equation

$$G_A\left(\frac{\rho(x, A)}{\rho(y, A)}\right) = G_B\left(\frac{\rho(x, B)}{\rho(y, B)}\right) \text{ whenever } x, y \in A \cap B, \tag{1}$$

where, for any $C \in \mathcal{A}$, the map G_C takes $(0, \infty)$ to $(0, \infty)$ and is strictly increasing.

A salient property of IIA is that it imposes restrictions only on behavior across, and not within, choice sets. In other words, the collection of choice rules obeying IIA is unconstrained in the sense that, after observing any (non-zero) choice probabilities on a single choice set, one can fill in the choice probabilities on other sets in a way that produces a choice rule obeying IIA. To state this formally, for any finite set Z, let ΔZ denote the set of all full-support probability measures on Z.

Definition 3. Let $P \subseteq \mathcal{P}$ be a collection of random choice rules. We say P is **free** if for every $A \in \mathcal{A}$ and $\lambda \in \Delta A$ there exists $\rho \in P$ with $\rho(\cdot, A) = \lambda$.

We want our relaxation of IIA to work similarly in imposing behavioral restrictions across and not within choice sets.

Definition 4. A family of functions $\{G_A\}_{A\in\mathcal{A}}$ mapping $(0,\infty)$ to $(0,\infty)$ is **admissible** if the G_A are strictly increasing and the collection of random choice rules that satisfy Equation (1) for the family $\{G_A\}_{A\in\mathcal{A}}$ is free.

An example of an admissible family of functions is obtained by fixing any function $F : \mathcal{A} \to (0,\infty)$ and setting $G_A(t) = t^{F(A)}$ for all $A \in \mathcal{A}$. (This is a corollary of Lemma 2 in the Appendix.) An example of an inadmissible family is any family such that $G_A(t) = \alpha t$ and $G_B(t) = \beta t$ where $\alpha \neq \beta$ and A and B have at least three elements in common. A demonstration of the inadmissibility of this class of families can be found in Appendix E.

Definition 5. A random choice rule ρ is a **relaxed IIA** rule if there exists an admissible family of functions $\{G_A\}_{A\in\mathcal{A}}$ with respect to which ρ obeys Equation (1).

2.4 The Equivalence Result

The main result of the paper establishes a three-way equivalence uniting our three perspectives on rationally imprecise behavior.

Theorem 1. *For any random choice rule ρ the following are equivalent:*

1. *ρ has an information-processing formulation.*

2. *ρ has a neural-normalization implementation.*

3. *ρ has a relaxed IIA characterization.*

We give a sketch of the proof in the next section. Full details are in the Appendix. Theorem 1 unites three perspectives on rationally imprecise behavior. The information-processing formulation explains why mistakes arise, namely, because of the costliness of reducing stochasticity in choice. The neural implementation explains how these mistakes can arise via a mechanism understood to be operative in the brain. The choice-rule characterization says exactly what behavior can be observed. The unification of the three perspectives comes through the observed behavior. Whenever choice probabilities fit into one of the three models, they must fit into all three.

The same three-way equivalence as in Theorem 1 can be proved for the classic IIA choice rule and restrictions of the information-processing and normalization models that involve constant $F(\cdot)$. The equivalence between classic IIA and the information-processing model with a constant $F(\cdot)$ was established by Mattsson and Weibull (2002). The normalization model with a constant $F(\cdot)$ reduces to the classic Logit model, which McFadden (1978) proved is equivalent to classic IIA. Theorem 1 extends these already-known relationships by allowing for the inclusion of the empirically-important factor $F(\cdot)$.

We now turn to the uniqueness properties of our representations. In the proof of Theorem 1, it is demonstrated that a pair (v, F) is an information-processing representation of ρ if and only if it also serves as neural representation of ρ. Thus, for the purposes of the uniqueness result, we will simply say "(v, F) represents ρ" to mean that (v, F) serves as both kinds of representations. There is one case when we cannot establish any uniqueness properties for F, which is when all alternatives in a choice set are chosen with equal probability. For our next result, we rule this case out.

Definition 6. A random choice rule ρ is **non-uniform** if for every $A \in \mathcal{A}$ with $|A| \geq 2$ there exist $x, y \in A$ with $\rho(x, A) \neq \rho(y, A)$.

Theorem 2. *Suppose (v, F) represents a non-uniform random choice rule ρ. Then (v', F')*

also represents ρ if and only if there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that:

$$\begin{aligned} v(x) &= \alpha v'(x) + \beta \textit{ for all } x \in X, \\ F(A) &= \alpha F'(A) \ \textit{ whenever } |A| \geq 2. \end{aligned}$$

This says that in any representation (v, F), the function v is unique up to an affine transformation and the function F is unique up to a multiplicative scaling factor. Notice that the multiplicative factor in the affine transformation of v must be the same as the scaling factor for F. An implication is that, while the ratios $v(\cdot)/F(A)$ are not unique, the differences between these ratios, namely,

$$\frac{v(x)}{F(A)} - \frac{v(y)}{F(A)},$$

must be unique.

3 Sketch of the Equivalence Proof

The proof of Theorem 1 proceeds by establishing that all three parts are equivalent to the statement that there exist functions $v : X \to (0, \infty)$ and $F : \mathcal{A} \to (0, \infty)$ such that

$$\rho(x, A) = \frac{e^{v(x)/F(A)}}{\sum_{y \in A} e^{v(y)/F(A)}} \text{ for all } A \in \mathcal{A} \text{ and } x \in A. \tag{2}$$

In the case of the information-processing and neural-normalization models, the functions v and F in the above equation will be exactly the representation of ρ. This establishes the claim, made in the uniqueness discussion in the previous section, that (v, F) serves as an information-processing representation if and only if it serves as a neural-normalization representation. We now sketch the argument for how each model in turn is equivalent to Equation (2).

That the information-processing formulation is equivalent to Equation (2) follows the standard proof in thermodynamics that minimizing Helmholtz free energy[5] yields the Boltzmann distribution. That the neural-normalization implementation is equivalent to Equation (2) is a straightforward extension of the work done by McFadden (1978). Our final, and novel, argument establishes the equivalence between our relaxed IIA rule and Equation (2). To show that a random choice rule satisfying Equation (2) must also obey the relaxed IIA rule, we set $G_A(t) = t^{F(A)}$ for each $A \in \mathcal{A}$. Details of this direction can be found in the Appendix.

To prove that any relaxed IIA choice rule ρ must follow Equation (2), start by fixing an admissible family of functions $\{G_A\}_{A\in\mathcal{A}}$ that map $(0,\infty)$ to $(0,\infty)$. Fix, in addition, a choice set $A \in \mathcal{A}$ with $|A| \geq 3$. (Dealing with sets that have two alternatives requires a different technique, which we defer to the full proof found in the Appendix.) Define a function $H_A : (0,\infty) \to (0,\infty)$ by setting

$$H_A(t) = G_X^{-1}(G_A(t)).$$

Taking any random choice rule ρ satisfying Equation (1) with respect to $\{G_A\}_{A\in\mathcal{A}}$, we have that for any pair of alternatives $x, y \in A$,

$$H_A\left(\frac{\rho(x,A)}{\rho(y,A)}\right) = \frac{\rho(x,X)}{\rho(y,X)}. \tag{3}$$

Therefore, for any triplet $x, y, z \in A$, we have

$$H_A\left(\frac{\rho(x,A)}{\rho(y,A)}\right) \times H_A\left(\frac{\rho(y,A)}{\rho(z,A)}\right) = \frac{\rho(x,X)}{\rho(z,X)} = H_A\left(\frac{\rho(x,A)}{\rho(z,A)}\right).$$

We show in the Appendix that, provided the family of functions $\{G_A\}_{A\in\mathcal{A}}$ is admissible, we can choose ρ to match any full-support distribution on A. Specifically, for any numbers

[5] Mandl (1988).

$a, b \in (0, \infty)$, we can choose ρ so that the previous equation becomes

$$H_A(a) \times H_A(b) = H_A(ab). \tag{4}$$

The next step involves the Cauchy functional equation in the form: Fix a function $H : \mathbb{R} \to \mathbb{R}$ which is continuous at one point (or more). Then, if H satisfies Equation (4), there exists $\beta \in \mathbb{R}$ such that $H(t) = t^\beta$.[6] In our setting, the exponent β must be strictly positive, and from this we get a function $F : \mathcal{A} \to (0, \infty)$ such that

$$H_A(t) = t^{F(A)}.$$

Equation (3) then becomes

$$\frac{\rho(x, A)}{\rho(y, A)} = \left(\frac{\rho(x, X)}{\rho(y, X)}\right)^{1/F(A)}, \tag{5}$$

from which, by summing over all $x \in A$,

$$\frac{1}{\rho(y, A)} = \sum_{x \in A} \left(\frac{\rho(x, X)}{\rho(y, X)}\right)^{1/F(A)},$$

which we can invert to get

$$\rho(y, A) = \frac{\rho(y, X)^{1/F(A)}}{\sum_{x \in A} \rho(x, X)^{1/F(A)}}.$$

Finally, in order to make the previous equation become Equation (2) simply define $v : X \to (0, \infty)$ as

$$v(x) = \alpha + \ln \rho(x, X),$$

where α is a constant chosen to be large enough to ensure $v(\cdot)$ is positive. The exact value of α is irrelevant, since it factors out when transforming the expression for $\rho(y, A)$ into

[6] See Theorem 3, p.41, in Aczél (1966).

Equation (2).

4 Axiomatic Characterization

We now turn to our revealed-preference exercise where we provide two testable axioms which characterize our relaxed IIA rule. Our first axiom allows for an ordinal ranking of choice alternatives, while our second axiom captures the type of set dependence allowed in our relaxed IIA rule. Unlike our choice rule, which requires searching over all possible functions $\{G_A\}_{A\in\mathcal{A}}$, our axioms are closed-form expressions of the choice probabilities and therefore fully operational.

Axiom 1 (Order)**.** *Let* $A, B \in \mathcal{A}$ *and* $x, y \in A \cap B$*. Then* $\rho(x, A) \geq \rho(y, A)$ *if and only if* $\rho(x, B) \geq \rho(y, B)$*.*

In both the information-processing and neural-normalization models, the inequality $v(x) \geq v(y)$ will hold if and only if x is chosen weakly more often than y in every choice set. This necessitates Axiom 1. Likewise, in our relaxed IIA rule, the functions $\{G_A\}_{A\in\mathcal{A}}$ are strictly increasing, which implies Axiom 1. To see why this is so, note that if ρ is a relaxed IIA rule, then

$$G_A\left(\frac{\rho(x,A)}{\rho(y,A)}\right) \geq G_A\left(\frac{\rho(y,A)}{\rho(x,A)}\right) \text{ implies } G_B\left(\frac{\rho(x,B)}{\rho(y,B)}\right) \geq G_B\left(\frac{\rho(y,B)}{\rho(x,B)}\right).$$

To motivate our second axiom we will use the following fact, which follows immediately from Equation (5) in the previous section.

Fact 1. *A random choice rule* ρ *is a relaxed IIA rule if and only if there is a function* $F : \mathcal{A} \to (0, \infty)$ *such that for all* $A, B \in \mathcal{A}$ *and* $x, y \in A \cap B$*,*

$$\left(\frac{\rho(x,A)}{\rho(y,A)}\right)^{F(A)} = \left(\frac{\rho(x,B)}{\rho(y,B)}\right)^{F(B)}.$$

This says that, for a relaxed IIA rule, it is without loss of generality to restrict attention to families of power functions, that is, to families of functions $G_A(t) = t^{F(A)}$ for some $F : \mathcal{A} \to (0, \infty)$.[7]

Using Fact 1, we see that a necessary condition for representation by a relaxed IIA rule is the existence of a function $F : \mathcal{A} \to (0, \infty)$ such that for any $x, y \in A \cap B$,

$$F(A) \ln \left(\frac{\rho(x, A)}{\rho(y, A)} \right) = F(B) \ln \left(\frac{\rho(x, B)}{\rho(y, B)} \right).$$

The same equality holds for $x', y' \in A \cap B$, so we must have

$$F(B) F(A) \ln \left(\frac{\rho(x, A)}{\rho(y, A)} \right) \ln \left(\frac{\rho(x', B)}{\rho(y', B)} \right) = F(A) F(B) \ln \left(\frac{\rho(x', A)}{\rho(y', A)} \right) \ln \left(\frac{\rho(x, B)}{\rho(y, B)} \right).$$

Canceling the term $F(A) \times F(B)$ from both sides yields our second axiom.

Axiom 2 (Log Ratio)**.** *For all $\{x, y, x', y'\} \subseteq A \cap B$,*

$$\ln \frac{\rho(x, A)}{\rho(y, A)} \ln \frac{\rho(x', B)}{\rho(y', B)} = \ln \frac{\rho(x', A)}{\rho(y', A)} \ln \frac{\rho(x, B)}{\rho(y, B)}. \tag{6}$$

Axiom 2 is a closed-form condition on the choice probabilities and is therefore directly testable. Specifically, the arbitrary function F with which we started is not involved in this formulation. We have just established that Axioms 1 and 2 are necessary for representation by a relaxed IIA rule. The next result, which is proved in the Appendix, says that they are also sufficient.

Theorem 3. *A random choice rule ρ satisfies relaxed IIA if and only if it obeys Axioms 1 and 2.*

Our axioms are independent in that neither implies the other. Consider that Axiom 1 implies that, in Equation (6), the two terms involving x and y must have the same sign.

[7] However, it is should be noted that not every admissible family $\{G_A\}_{A \in \mathcal{A}}$ is a family of power functions. An example is provided in the Appendix.

However, this is not an implication of Axiom 2 alone. Axiom 2 could be satisfied if, for example, in the terms in Equation (6) involving set A were both positive and the terms involving set B were both negative. Therefore, Axiom 2 does not imply Axiom 1. The appendix contains a fully worked-out example demonstrating this fact. That Axiom 1 does not imply Axiom 2 is clear from inspection.

Axiom 2 appears previously in Marley, Flynn, and Louviere (2008), who use it axiomatize a model defined by Equation (2) and therefore equivalent to ours. However, the independence of our two axioms implies both of them are necessary for Theorem 1, and hence the result in Marley, Flynn, and Louviere (2008, Theorem 3) does not go through as stated.[8]

5 Behavioral Properties

Our axioms in Section 4 fully characterize our relaxed IIA rule, but they do not make explicit the relationship to such behavioral properties as regularity and stochastic transitivity. (We define the latter, in weak and strong forms, below.) In this section, we show that, consistent with evidence from the literature on the attraction effect, our relaxed IIA rule permits violations of regularity. We also show that our choice rule must satisfy the weak but not the strong form of stochastic transitivity. This, too, seems broadly in line with the empirical evidence (Rieskamp, Busemeyer, and Mellers, 2006). Lastly, we provide a novel testable prediction on the effect of the choice set on stochasticity of choice — a prediction which suggests a direction for future empirical work.

Ordinality and Stochastic Transitivity

From Axiom 1 we can see that our relaxed IIA rule satisfies **ordinality** in the sense that whether one alternative is more or less likely to be chosen over another alternative is answered the same way across different choice sets. In terms of a representation (v, F), this relationship

[8]This point became clear after a fruitful discussion with Tony Marley, to whom we are very grateful.

is expressed through the value function v: If $v(x)$ is greater then $v(y)$, then $\rho(x, A)$ is greater than $\rho(y, A)$ whenever $x, y \in A$. This is immediate from Equation (2) in Section 4.

Closely related to ordinality is the notion of **stochastic transitivity** (**ST**), which comes in two forms. Weak ST says that if $\rho(x, \{x, y\}) \geq \frac{1}{2}$ and $\rho(y, \{y, z\}) \geq \frac{1}{2}$, then $\rho(x, \{x, z\}) \geq \frac{1}{2}$. Strong ST says that if $\rho(x, \{x, y\}) \geq \frac{1}{2}$ and $\rho(y, \{y, z\}) \geq \frac{1}{2}$, then $\rho(x, \{x, z\}) \geq \max\{\rho(x, \{x, y\}), \rho(y, \{y, z\})\}$. It is easily checked that the classic IIA rule satisfies both forms of ST. Rieskamp, Busemeyer, and Mellers (2006) survey the empirical evidence and observe that while violations of weak ST are rare and occur only in special circumstances, violations of strong ST are frequent and occur in a wide variety of circumstances. They conclude that reasonable models of choice should satisfy ST in its weak but not in its strong form.

Consistent with this conclusion, our relaxed IIA rule satisfies weak ST but can violate strong ST. Satisfaction of weak ST is readily checked via Equation (2). The following example shows how strong ST can be violated. The key to generating a violation is to choose the set-dependent scaling $F(\{x, z\})$ (relative to the scalings $F(\{x, y\})$ and $F(\{y, z\})$) so as to reduce the difference between the values $v(x)$ and $v(z)$ (relative to the differences between $v(x)$ and $v(y)$, and between $v(y)$ and $v(z)$).

Example 1. Let $X = \{x, y, z\}$. Let (v, F) be a representation with $v(x) = 3, v(y) = 2, v(z) = 1$, and $F(\{x, y\}) = F(\{y, z\}) = 1$, $F(\{x, z\}) = 10$. From Equation (2) we get $\rho(x, \{x, y\}) = \rho(y, \{y, z\}) \approx 0.731$ and $\rho(x, \{x, z\}) \approx 0.550$. Thus $\rho(x, \{x, y\})$ and $\rho(y, \{y, z\})$ are both greater than $\frac{1}{2}$ and also greater than $\rho(x, \{x, z\})$, violating strong ST.

Regularity

A random choice rule ρ obeys **regularity** if $x \in A \subseteq B$ implies $\rho(x, A) \geq \rho(x, B)$. A random utility model satisfies regularity, as do previous generalizations of the classic IIA rule such as ordinal IIA (Fudenberg, Iijima, and Strzalecki, 2015) and the attribute rule (Gul, Natenzon, and Pesendorfer, 2014). However, violations of regularity have been widely documented, specifically, in the well-known attraction effect (Huber, Payne, and Puto, 1982; Simonson, 1989).

The set-dependent way in which we relax IIA allows our choice rule to accommodate violations of regularity. In our representational models, the set dependence is expressed through the function F. In the information-processing case, a higher F increases the cost of reducing randomness, which therefore favors inferior alternatives. Likewise, in the neural-normalization case, a higher F re-scales the values of alternatives to be closer together, which again favors inferior alternatives. Thus, by varying F across choice sets in the right way, we can add an alternative to a choice set so that the probability of choosing an existing alternative increases, which is a violation of regularity. The following example illustrates this effect.

Example 2. Let X=$\{x, y, z\}$. Fix a representation (v, F) with $v(x) = 3, v(y) = 2, v(z) = 1$, and $F(\{x, z\}) = 1$, $F(\{x, y, z\}) = 10$. From Equation (2) we get $\rho(z, \{x, z\}) \approx 0.119$ and $\rho(z, \{x, y, z\}) \approx 0.301$, violating regularity.

This example is not a knife-edge or special case. To show this, we consider the question: Suppose we observe the choice probabilities from a set A. Under what conditions does a relaxed IIA rule exist that matches the observed probabilities on A, and that creates a regularity violation when adding or removing an item from A? The next proposition (which is proved in the Appendix) offers such conditions, and they can be seen to be quite unrestrictive.

Proposition 1. *Fix a choice set A and some $\lambda \in \Delta A$. Then:*

1. *For any $x \in A$ and $y \notin A$ there is a relaxed IIA rule ρ such that: (i) $\rho(\cdot, A) = \lambda$; and (ii) $\rho(x, A) < \rho(x, A \cup \{y\})$ if and only if λ is non-uniform on A.*

2. *For any $x, y \in A$ with $x \neq y$, there is a relaxed IIA rule ρ such that: (i) $\rho(\cdot, A) = \lambda$; and (ii) $\rho(x, A \setminus \{y\}) < \rho(x, A)$ if and only if $\lambda(x) < \max_{z \in A \setminus \{y\}} \lambda(z)$ or $\lambda(x) > \frac{1}{|A|-1}$.*

The ability of the relaxed IIA rule to accommodate regularity violations is not without limits. For example, any regularity violation which fails Axiom 1 (Order) cannot arise from a relaxed IIA rule. At first blush, evidence from the literature on the attraction effect (Huber, Payne, and Puto, 1982; Simonson, 1989) would seem to cast doubt on the empirical validity of our Axiom 1. But, most of the relevant studies involve aggregating data across subjects. This is important because it is entirely possible for aggregate data to violate our Axiom 1 even when each individual satisfies it, as the next example shows. (We note that this is not true for regularity. If each individual satisfies regularity, then so will the aggregate data.) Studies that do perform within-subject analysis (e.g., Berkowitsch, Scheibehenne, and Rieskamp, 2014) often calibrate the choices to be very close to indifference in one of the choice sets, which makes it difficult to ascertain whether or not violations of Axiom 1 pass statistical significance.

Example 3. There are two types of individuals in a population, with equal numbers of type I and type II. When faced with a choice between alternatives x and y, type I individuals choose x with probability 0.8, while type II individuals choose y with probability 0.7. There is also a third option z (it would be called a 'decoy' in the attraction-effect literature) which, if present, makes option y look more appealing. When choosing from the set $\{x, y, z\}$, type I individuals choose x with probability 0.6, y with probability 0.3, and z with probability 0.1. In the same scenario, type II individuals choose x with probability 0.1, y with probability 0.8, and z with probability 0.1. It is clear that, separately, both types obey Axiom 1. But it is also clear that the aggregate data do not. From set $\{x, y\}$ the total population chooses x with probability 0.55 and y with probability 0.45, but from set $\{x, y, z\}$ the population chooses x with probability 0.35 and y with probability 0.55.

Relative Stochasticity

A novel prediction of our relaxed IIA rule is that choice sets can be ranked by how they affect the stochasticity of choice. We will think of choices as more stochastic in a first choice set as compared with a second choice set if alternatives in common across the two sets are closer to equiprobable in the first set. Of course, comparing choice sets in this way can be be done only when the sets share at least two alternatives. (This condition will be assumed below.)

Definition 7. A random choice rule ρ is **more stochastic** on choice set A than on choice set B if for every $x, y \in A \cap B$ with $\rho(x, A) \geq \rho(y, A)$, we have

$$\frac{\rho(x, A)}{\rho(y, A)} \leq \frac{\rho(x, B)}{\rho(y, B)}.$$

Of course, we always get equality here under the classic IIA rule. That is, a classic IIA rule is equi-stochastic on any two choice sets. At the other extreme, it is straightforward to construct examples, involving a general random choice rule, where the choice probabilities of two alternatives x and y are closer to equiprobable in set A as compared with set B, but the reverse is true for two other alternatives x' and y'. In this case, we cannot say whether the rule is more or less stochastic across the two sets. Our choice rule ρ occupies a middle ground, in that, for a given such rule ρ and pair of choice sets (with two or more elements in common), there is an unambiguous ranking of stochasticity across the two sets.

Proposition 2. *Let ρ be a relaxed IIA rule. Then, for any two choice sets A and B, either ρ is more stochastic on A than on B, or ρ is more stochastic on B than on A.*

Turning to (v, F) representations of our relaxed IIA rule, we can link the stochasticity relationship directly to the function $F(\cdot)$. A larger value for $F(\cdot)$ increases stochasticity, while a smaller value decreases stochasticity. More precisely, a relaxed IIA rule ρ is more stochastic on A than on B if and only if $F(A) \geq F(B)$. (The proofs of Proposition 2 and this statement can be found in the Appendix.)

As far as we are aware, the prediction that there is an unambiguous ranking of stochasticity across choice sets is new. This suggests a new area for empirical work.

6 Concluding Remarks

In this paper, we examined mistaken-prone or rationally imprecise behavior through three separate questions. We first asked why such behavior might occur, that is, what factors and what trade-offs explain the occurrence of mistakes. We next asked how the behavior might be implemented in the human brain. Last, we asked what choice rule characterizes the behavior. Our answers to these three questions came, respectively, in the form of our information-processing formulation, our neural-normalization implementation, and our relaxed IIA choice rule. These three seemingly disparate models become united through our equivalence result, which demonstrates that each model provides a different perspective on the same behavior.

Our information-processing formulation of rational imprecision sits at the intersection of economics and neuroscience. Not surprisingly, information-theoretic methods have featured in both fields. In economics, the rational-inattention model of Sims (2003) uses the cost of information acquisition to explain why a rational decision maker might fail to use all available signals. In neuroscience, the notion of efficient coding of information has played a key role in understanding the structure of sensory systems (Schwartz and Simoncelli, 2001). Likewise, we take an information-theoretic approach — in our case, by taking into account the information-processing cost of reducing stochasticity in choice behavior to explain why a rational decision maker might make mistakes.

Our neural-normalization implementation seeks to ground rationally imprecise behavior in a neural architecture that can generate it. This grounding is done through the normalization computation that has been successful at explaining the neural activity associated with value representation (Louie, Grattan, and Glimcher, 2011). More recently, normalization has also been used to explain choice behaviors including many of those associated with prospect theory (see Footnote 2).

Our relaxed IIA choice rule characterizes rationally imprecise behavior, which allows us both to examine how our model can match empirically observed choice behavior and to generate new testable predictions. The relaxed IIA rule can accommodate violations of regularity, including the attraction effect, which is a point of contrast with standard models such as random utility. Additionally, the relaxed IIA rule allows violations of strong, but not weak, stochastic transitivity, which is broadly consistent with observed behavior. We generate closed-form axioms equivalent to our relaxed IIA rule which provide testable predictions. We also provide a novel prediction that choice sets can be ranked by how they affect the stochasticity of choice.

We conclude by commenting on the role of the most unusual aspect, relative to the economics literature, of our three-pronged approach — namely, our neural implementation argument. The inclusion of this aspect is motivated, in part, by the argument due to Simon (1955, p. 99) that a theory of decision making should be consistent "with the access to information and the computational capacities that are actually possessed by the organism." At the time of Simon's writing, the development of such a theory was hindered by a lack of empirical knowledge about precisely such information and computational capacities — a fact which Simon himself noted (Simon 1955, p. 100).

In the decades since then, advances in neuroscience have taught us a lot about the actual decisional processes of various organisms, humans included. By capitalizing on these advances, we have been able to build a theory of decision-making consistent with how the human brain actually makes choices, and, in this way, advance Simon's argument. With this, we hope to have taken a step towards reconciling traditional approaches to decision-making with the fact that all behavior making must, ultimately, have a physical implementation.

Appendix

A Background on the Normalization Computation

The neural-implementation model in this paper is based on the canonical normalization computation identified by neurobiologists over the past two decades. This appendix is intended to supply the interested reader with a brief overview of the origins and ubiquity of the normalization forms in the neuroscientific literature. A more detailed review can be found in Glimcher (2014).

Inspired by earlier arguments of Barlow (1961) and Attneave (1954) that neural processes should be efficiency-promoting, Heeger, Simoncelli, and Movshon (1996) proposed a normalization computation in which the visual stimulus observed at one point in the visual field is re-scaled in a way that depends on other points in the field. Thus, letting X_i, for $i = 1, ..., n$, be the luminance of point i in the set constituting the currently viewed image, normalization computes the firing rate of the neuron responsible for encoding X_i to be

$$FR_i := \frac{X_i}{\sigma + \sum_{j=1}^{n} \omega_{i,j} X_j} + \varepsilon_i. \tag{7}$$

Here, the $\omega_{i,j}$'s are fixed weights indicating how much each point j should influence the computation for point i, the constant σ is a fixed parameter, and ε_i is a stochastic error term that describes noise in the visual system.

Experimental biologists brought this functional form to the study of the parts of the mammalian brain that are involved in the representation of visual stimuli. They found, almost without exception, that the normalization formula was more successful than previous functional forms in modeling the visual representation function. Subsequent research demonstrated that essentially all parts of the mammalian brain that represent external states of the world — what a neuroscientist calls "sensory representations" — are successfully described by normalization-like functional forms. Carandini and Heeger (2012) is a useful survey of the evidence supporting the presence of normalization in the various sensory domains.

Around the same time, a separate literature identified in primates a discrete set of brain areas apparently dedicated to representing a quantity that functions much like an economic utility representation (e.g., Platt and Glimcher 1999; Dorris and Glimcher 2004; Barraclough, Conroy, and Lee 2004). Specifically, these brain areas assign different levels of activity to different choice alternatives, where higher associated activity indicates a higher probability that the alternative in question is chosen. Subsequent fMRI brain-scanning studies demonstrated that these same representations arise in human brains (e.g., Knutson et al. 2001; Delgado et al. 2000; Elliott et al. 2000; Kable and Glimcher 2007).

Given the ubiquity of normalization computations in the sensory systems, one might hypothesize that utility representation in the brain also operates via normalization. Relative to Equation (7), normalization in the choice domain replaces luminance and the visual field with the value of each alternative and the choice set, respectively. Each alternative y is assigned a value $v(y)$, and the utility of alternative x in choice set A is given by

$$U_x := \frac{v(x)}{\sigma + \sum_{y \in A} v(y)} + \varepsilon_x. \tag{8}$$

A direct study of brain activity in primates by Louie, Grattan, and Glimcher (2011) (subsequently, Louie et al. (2014) and Hunt, Dolan, and Behrens (2014)) showed that Equation (8) is effective in describing the utility-like mechanisms in the brain. A utility function following Equation (8) has even been successful in explaining human behavior in choice experiments (e.g., Louie, Khaw, and Glimcher 2013).

The neural-normalization implementation in the present paper generalizes Equation (8) by allowing the denominator to be any function of the choice set. Our generalization is aimed, in part, at allowing analogs to the weights in the visual normalization formula (Equation 7), since these are missing from the value normalization formula (Equation 8). We also make specific distribution assumptions on the error term ε_x.

The preceding discussion implicitly assumes that neural studies on primates and other mammals are relevant and insightful in the case of human neural architecture. This assumption is justified by the remarkable consistency with which, over the past 150 years, studies of various neural systems in animals have reliably revealed the structures of the analogous systems in humans.[9] Of course, this consistency is, at bottom, a consequence of evolution. But, importantly, it has been confirmed in repeated specific cases. Thus, thirty years ago, studies of spatial navigation in rodents revealed the mechanism of human spatial navigation (see Best, White, and Minai 2001). Twenty years ago, studies of visual and acoustic processing centers in monkeys revealed the foundations of human perceptual experience (Hubel and Wiesel, 2004). More recently, studies of social interactions in monkeys have revealed the biological foundations of trust and empathy in humans (Bernhardt and Singer, 2012).

[9] Kandel, Schwartz, and Jessel (2000) is a standard reference.

B Proof of Theorem 1

As in the proof sketch in Section 3, we prove that each part of Theorem 1 is equivalent to the statement that ρ has a representation in the form of Equation (2).

Proof for Theorem 1.1: We will solve a relaxed version of the maximization problem, which allows choice probabilities to be 0. But we will find that the solution does not involve setting any probabilities equal to 0, so that this also solves the actual maximization problem.

Suppose ρ has an information-processing formulation (v, F), so that, for every $A \in \mathcal{A}$, the rule ρ maximizes

$$\sum_{x \in A} \left\{ \rho(x, A) v(x) - F(A) \rho(x, A) \ln(\rho(x, A)) \right\}$$

subject to

$$\begin{aligned} \sum_{x \in A} \rho(x, A) &= 1, \\ \rho(x, A) &\geq 0. \end{aligned}$$

By the Kuhn-Tucker theorem, the necessary first-order condition is that there exist $\lambda \in \mathbb{R}$ and $\eta_x \in \mathbb{R}$, for every $x \in X$, such that for each $\rho(A, x)$ we have

$$v(x) - F(A)(\ln(\rho(x, A)) + 1) + \lambda + \eta_x = 0.$$

Notice that $\lim_{x \to 0} \ln(x) = -\infty$, so the first-order condition can never be satisfied when $\rho(A, x) = 0$. (This can also be understood in terms of the behavior of the entropy. As $\rho(A, x) \to 0$, the derivative of $H(\rho(A, \cdot))$ with respect to $\rho(A, x)$ goes to $-\infty$, even though H is itself uniformly bounded.) Thus, it is never optimal to set $\rho(A, x) = 0$. By Complementary Slackness, it follows that $\eta_x = 0$ for all $x \in A$ and we can solve the first-order condition to get

$$\rho(x, A) = \exp\left(\frac{v(x) + \lambda}{F(A)} - 1\right).$$

We can then divide through to get

$$\frac{\rho(x, A)}{\rho(y, A)} = \exp\left(\frac{v(x) - v(y)}{F(A)}\right).$$

Next note that

$$\sum_{x\in A}\frac{\rho\left(x,A\right)}{\rho\left(y,A\right)}=\frac{\sum_{x\in A}\rho\left(A,x\right)}{\rho\left(y,A\right)}=\frac{1}{\rho\left(y,A\right)}=\frac{\sum_{x\in A}\exp\left(\frac{v(x)}{F(A)}\right)}{\exp\left(\frac{v(y)}{F(A)}\right)}$$

which implies

$$\rho\left(y,A\right)=\frac{\exp\left(\frac{v(y)}{F(A)}\right)}{\sum_{x\in A}\exp\left(\frac{v(x)}{F(A)}\right)},$$

establishing that ρ obeys Equation (2).

To prove the other direction, notice that Equation (2) is both a necessary condition for the information-processing problem and fully pins down the solution. Moreover, by the Weierstrass theorem, the maximization problem has at least one solution.

Proof for Theorem 1.2: Proving the equivalence of our neural-normalization implementation and Equation (2) follows the lines of well-known arguments in McFadden (1978). To begin, suppose ρ has a neural-normalization implementation (v,F), so that

$$\rho\left(x,A\right)=\Pr\left(x=\arg\max\frac{v\left(y\right)}{F\left(A\right)}+\varepsilon_y\right),$$

where ε_y's are i.i.d. and Gumbel with location 0 and scale 1. Let $g(t)=\exp\left(-t-\exp\left(-t\right)\right)$ and $G(t)=\exp\left(-\exp\left(-t\right)\right)$ be the pdf and cdf of a Gumbel $(0,1)$ random variable. We then have

$$\rho\left(x,A\right)=\int_{-\infty}^{+\infty}\left\{\prod_{y\in A\setminus\{x\}}G\left(\frac{v\left(x\right)-v\left(y\right)}{F\left(A\right)}+t\right)\right\}g\left(t\right)dt=$$
$$\int_{-\infty}^{+\infty}\left\{\prod_{y\in A\setminus\{x\}}\exp\left(-\exp\left(-\frac{v\left(x\right)-v\left(y\right)}{F\left(A\right)}-t\right)\right)\right\}\exp\left(-t-\exp\left(-t\right)\right)dt,$$

which we can rearrange to give

$$\rho\left(x,A\right)=\int_{-\infty}^{+\infty}\left\{\exp\left(-\exp\left(-t\right)\left(1+\sum_{y\in A\setminus\{x\}}\exp\left(-\frac{v\left(x\right)-v\left(y\right)}{F\left(A\right)}\right)\right)\right)\right\}\exp\left(-t\right)dt,$$

which can be integrated to obtain

$$\rho(x,A) = \frac{1}{\left(1+\sum_{y\in A\setminus\{x\}}\exp\left(-\frac{v(x)-v(y)}{F(A)}\right)\right)}\times$$
$$\exp\left(-\exp(-t)\left(1+\sum_{y\in A\setminus\{x\}}\exp\left(-\frac{v(x)-v(y)}{F(A)}\right)\right)\right)\Bigg|_{t=-\infty}^{t=+\infty}.$$

Evaluating at the limits yields

$$\rho(x,A) = \frac{1}{1+\sum_{y\in A\setminus\{x\}}\exp\left(-\frac{v(x)-v(y)}{F(A)}\right)}(1-0),$$

which can be rearranged to give

$$\rho(x,A) = \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\sum_{y\in A}\exp\left(\frac{v(y)}{F(A)}\right)},$$

as desired. This argument can be run backwards to prove the reverse implication.

Proof for Theorem 1.3: Suppose that ρ obeys Equation (2) for some $v : X \to (0,\infty)$ and $F : \mathcal{A} \to (0,\infty)$. Define $G_A : (0,\infty) \to (0,\infty)$ by setting $G_A(t) = t^{F(A)}$. Lemma 2 (established later in the Appendix) tells us that $\{G_A\}_{A\in\mathcal{A}}$ is admissible. Now, for $A \in \mathcal{A}$ and $x, y \in A$, we can write

$$G_A\left(\frac{\rho(x,A)}{\rho(y,A)}\right) = G_A\left(\exp\left(\frac{v(x)-v(y)}{F(A)}\right)\right) = \exp(v(x)-v(y)),$$

and, since the right-hand side is independent of A, we see that ρ obeys Equation (1) and is therefore a relaxed IIA rule.

For the reverse direction, we prove that if ρ is a relaxed IIA rule, then there exists $F : \mathcal{A} \to (0,\infty)$ such that $F(X) = 1$ and ρ obeys Equation (1) where $G_A(t) = t^{F(A)}$. Proving this claim will be sufficient to establish that ρ obeys Equation (2). To see this, note that for any $A \in \mathcal{A}$ and $x, y \in A$, we then get

$$\frac{\rho(y,A)}{\rho(x,A)} = \left(\frac{\rho(y,X)}{\rho(x,X)}\right)^{1/F(A)}.$$

Summing over $y \in A$ yields

$$\frac{1}{\rho(x,A)} = \frac{\sum_{y\in A}\rho(y,X)^{1/F(A)}}{\rho(x,X)^{1/F(A)}},$$

from which,

$$\rho(x,A) = \frac{\rho(x,X)^{1/F(A)}}{\sum_{y\in A}\rho(y,X)^{1/F(A)}}.$$

Next define $v : X \to (0,\infty)$ by

$$v(x) = \alpha + \ln\rho(x,X)$$

where

$$\alpha := 1 - \min_{y\in X}\rho(y,X)$$

so that we get

$$\rho(x,A) = \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\sum_{y\in A}\exp\left(\frac{v(y)}{F(A)}\right)},$$

as desired.

We now turn to the proof of our claim. We need a preliminary step.

Lemma 1. *For any $|A| \geq 3$ and $a,b \in (0,\infty)$, there exist $\lambda \in \Delta A$ and $x,y,z \in A$ such that $\lambda(x)/\lambda(y) = a$ and $\lambda(y)/\lambda(z) = b$.*

Proof. If $|A| = 3$ set $c = 1$, otherwise set $c = \frac{1}{2}$. Let x,y,z be three distinct elements in A. Now define $\lambda \in \Delta A$ by

$$\begin{aligned}\lambda(x) &= \frac{cba}{1+b+ba},\\ \lambda(y) &= \frac{cb}{1+b+ba},\\ \lambda(z) &= \frac{c}{1+b+ba},\end{aligned}$$

and, if $|A| > 3$, by $\lambda(w) = \frac{1}{2(|A|-3)}$ for all $w \notin \{x,y,z\}$. It is easy to see that that λ has full support and that $\lambda(x)/\lambda(y) = a$ and $\lambda(y)/\lambda(z) = b$. Finally, if $|A| = 3$, then

$$\sum_{w\in A}\lambda(w) = \frac{cba+cb+c}{1+b+ba} = c = 1,$$

and if $|A| > 3$, then

$$\sum_{w\in A}\lambda(w) = \sum_{w\in A\setminus\{x,y,z\}}\frac{1}{2(|A|-3)} + \frac{cba+cb+c}{1+b+ba} = \frac{1}{2} + c = 1.$$

Thus λ is a well-defined element of ΔA. $\square$

We now proceed with the proof of our claim. Let ρ be a relaxed IIA rule, so that there is an admissible family $\{G_A\}_{A\in\mathcal{A}}$ under which ρ obeys Equation (1). We suppose that $|X| \geq 3$. (The cases $|X| = 2$ and $|X| = 1$ are trivial and omitted.) Since G_A is strictly increasing, it is invertible. For any $A \in \mathcal{A}$, define $H_A : (0,\infty) \to (0,\infty)$ by

$$H_A(a) = G_X^{-1}(G_A(a)).$$

Fix any $A \in \mathcal{A}$ with $|A| \geq 3$. By Lemma 1, for any $a, b \in (0,\infty)$, we can find three distinct elements $x, y, z \in A$, and $\lambda \in \Delta A$, such that

$$\frac{\lambda(x)}{\lambda(y)} = a \text{ and } \frac{\lambda(y)}{\lambda(z)} = b.$$

Since $\{G_A\}_{A\in\mathcal{A}}$ is admissible there exists a random choice rule $\tilde{\rho}$ which obeys Equation (1) under $\{G_A\}_{A\in\mathcal{A}}$ and satisfies $\tilde{\rho}(\cdot, A) = \lambda$. We have that for any $x', y' \in \{x, y, z\}$,

$$G_A\left(\frac{\tilde{\rho}(x', A)}{\tilde{\rho}(y', A)}\right) = G_X\left(\frac{\tilde{\rho}(x', X)}{\tilde{\rho}(y', X)}\right),$$

which implies that

$$H_A\left(\frac{\tilde{\rho}(x', A)}{\tilde{\rho}(y', A)}\right) = \frac{\tilde{\rho}(x', X)}{\tilde{\rho}(y', X)}.$$

From this we get

$$H_A\left(\frac{\tilde{\rho}(x, A)}{\tilde{\rho}(y, A)}\right) \times H_A\left(\frac{\tilde{\rho}(y, A)}{\tilde{\rho}(z, A)}\right) = \frac{\tilde{\rho}(x, X)}{\tilde{\rho}(y, X)} \times \frac{\tilde{\rho}(y, X)}{\tilde{\rho}(z, X)} = \frac{\tilde{\rho}(x, X)}{\tilde{\rho}(z, X)},$$

and also

$$H_A\left(\frac{\tilde{\rho}(x, A)}{\tilde{\rho}(z, A)}\right) = \frac{\tilde{\rho}(x, X)}{\tilde{\rho}(z, X)}.$$

Combining the previous two equations yields

$$H_A\left(\frac{\tilde{\rho}(x, A)}{\tilde{\rho}(y, A)}\right) \times H_A\left(\frac{\tilde{\rho}(y, A)}{\tilde{\rho}(z, A)}\right) = H_A\left(\frac{\tilde{\rho}(x, A)}{\tilde{\rho}(z, A)}\right),$$

from which, using $\tilde{\rho}(A, \cdot) = \lambda$, we get

$$H_A(a) \times H_A(b) = H_A(ab)$$

for any $a, b \in (0,\infty)$. Moreover, the functions G_A and G_X are increasing and therefore have at most a countable number of discontinuities. Thus, the map H_A can have at most a countable number of discontinuities. By a version of the Cauchy functional theorem (refer back to the proof sketch in Section 3), there exists $\tilde{F} : \mathcal{A}^{\geq 3} \to \mathbb{R}$, where $\mathcal{A}^{\geq 3}$ is the set of all $A \in \mathcal{A}$ with $|A| \geq 3$, such that

$$H_A(t) = t^{\tilde{F}(A)}.$$

Clearly $\tilde{F}(X) = 1$. Recalling the definition of H_A, we get

$$G_A(t) = G_X(t^{\tilde{F}(A)}).$$

Since G_A and G_X are both strictly increasing, it follows that $\tilde{F}(A) > 0$ for all $A \in \mathcal{A}^{\geq 3}$. We can also write

$$G_A(t^{\frac{1}{\tilde{F}(A)}}) = G_X(t).$$

Thus, for any $A, B \in \mathcal{A}^{\geq 3}$,

$$G_A(t^{\frac{1}{\tilde{F}(A)}}) = G_B(t^{\frac{1}{\tilde{F}(B)}}),$$

from which,

$$G_A(t) = G_B(t^{\frac{\tilde{F}(A)}{\tilde{F}(B)}}).$$

Since G_A, G_B are strictly increasing, we know that $G_A(t) = G_B(u)$ if and only if $u = t^{\frac{\tilde{F}(A)}{\tilde{F}(B)}}$, which is true if and only if $t^{\tilde{F}(A)} = u^{\tilde{F}(B)}$. Since ρ obeys Equation (1) under $\{G_A\}_{A\in\mathcal{A}}$, we conclude that for any $A, B \in \mathcal{A}^{\geq 3}$ and $x, y \in A \cap B$,

$$\left(\frac{\rho(x, A)}{\rho(y, A)}\right)^{\tilde{F}(A)} = \left(\frac{\rho(x, B)}{\rho(y, B)}\right)^{\tilde{F}(B)},$$

as desired.

We still have to deal with the case $|A| = 2$. Set $A = \{x, y\}$ and let $\alpha_A \in \mathbb{R}$ solve

$$\left(\frac{\rho(x, A)}{\rho(y, A)}\right)^{\alpha_A} = \left(\frac{\rho(x, X)}{\rho(y, X)}\right)^{\tilde{F}(X)} = \frac{\rho(x, X)}{\rho(y, X)}.$$

Since G_A is strictly increasing, we have $\frac{\rho(x,A)}{\rho(y,A)} \geq 1$ if and only if $G_A\left(\frac{\rho(x,A)}{\rho(y,A)}\right) \geq G_A\left(\frac{\rho(y,A)}{\rho(x,A)}\right)$. Since ρ obeys Equation (1) under $\{G_A\}_{A\in\mathcal{A}}$, we get $\frac{\rho(x,A)}{\rho(y,A)} \geq 1$ if and only if $G_X\left(\frac{\rho(x,A)}{\rho(y,A)}\right) \geq G_X\left(\frac{\rho(y,A)}{\rho(x,A)}\right)$. Since G_X is strictly increasing, we then have $\frac{\rho(x,A)}{\rho(y,A)} \geq 1$ if and only if $\frac{\rho(x,X)}{\rho(y,X)} \geq 1$. It follows that we can choose $\alpha_A \in (0, \infty)$.

Now define $F : \mathcal{A} \to (0, \infty)$ by

$$F(A) = \begin{cases} \tilde{F}(A) & \text{if} \quad |A| \geq 3, \\ \alpha_A & \text{if} \quad |A| = 2, \\ 1 & \text{if} \quad |A| = 1. \end{cases}$$

It can easily be verified that for all $A, B \in \mathcal{A}$ and $x, y \in A \cap B$, we have

$$\left(\frac{\rho(y, A)}{\rho(x, A)} \right)^{F(A)} = \left(\frac{\rho(y, B)}{\rho(x, B)} \right)^{F(B)},$$

as desired.

C Proof of Theorem 2

Suppose ρ is non-uniform and has representations (v, F) and (v', F'). Since ρ is non-uniform there exists $\bar{\mathbf{x}}, \underline{\mathbf{x}}$ such that $\rho(\bar{\mathbf{x}}, X) > \rho(\underline{\mathbf{x}}, X)$. Equation (2) implies that for any $y \in X$,

$$\frac{v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}})}{v(\bar{\mathbf{x}}) - v(y)} = \left(\ln \frac{\rho(\bar{\mathbf{x}}, X)}{\rho(\underline{\mathbf{x}}, X)} \right) \left(\ln \frac{\rho(\bar{\mathbf{x}}, X)}{\rho(y, X)} \right)^{-1} = \frac{v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}})}{v'(\bar{\mathbf{x}}) - v'(y)}.$$

Rearranging gives

$$(v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}}))(v'(\bar{\mathbf{x}}) - v'(y)) = (v(\bar{\mathbf{x}}) - v(y))(v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}})),$$

which can be further rearranged to give

$$v(y) = v(\bar{\mathbf{x}}) - \frac{(v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}}))}{(v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}}))}(v'(\bar{\mathbf{x}}) - v'(y)),$$

from which, finally,

$$v(y) = v'(y) \frac{v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}})}{v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}})} + v(\bar{\mathbf{x}}) - \frac{v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}})}{v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}})} v'(\bar{\mathbf{x}}).$$

Now set

$$\alpha = \frac{v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}})}{v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}})} \text{ and } \beta = v(\bar{\mathbf{x}}) - \frac{v(\bar{\mathbf{x}}) - v(\underline{\mathbf{x}})}{v'(\bar{\mathbf{x}}) - v'(\underline{\mathbf{x}})} v'(\bar{\mathbf{x}}).$$

Since $\rho(\bar{\mathbf{x}}, X) > \rho(\underline{\mathbf{x}}, X)$, it is clear that $\alpha > 0$. Fix any $A \in \mathcal{A}$ with $|A| \geq 2$. Since ρ is non-uniform, there exists $x, y \in A$ with $\rho(x, A) \neq \rho(y, A)$. We can write

$$\ln\left(\frac{\rho(x, A)}{\rho(y, A)}\right) = \frac{v(x) - v(y)}{F(A)} = \frac{v'(x) - v'(y)}{F'(A)},$$

from which, using the relationship between v and v', we get

$$\frac{\alpha v'(x) - \alpha v'(y)}{F(A)} = \frac{v'(x) - v'(y)}{F'(A)}.$$

Since $\rho(x, A) \neq \rho(y, A)$, we know that $v'(x) \neq v'(y)$, so we conclude which implies that $F(A) = \alpha F'(A)$, as desired.

For the reverse direction, suppose that (v, F) represents ρ and let $\alpha > 0$ and $\beta \in \mathbb{R}$. Equation (2) then yields, for every $A \in \mathcal{A}$,

$$\begin{aligned} \rho(x, A) &= \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\sum_{y \in A} \exp\left(\frac{v(y)}{F(A)}\right)} = \frac{\exp\left(\frac{\alpha v(x)}{\alpha F(A)}\right)}{\sum_{y \in A} \exp\left(\frac{\alpha v(y)}{\alpha F(A)}\right)} = \frac{\exp\left(\frac{\beta}{\alpha F(A)}\right)\left(\exp\left(\frac{\alpha v(x)}{\alpha F(A)}\right)\right)}{\exp\left(\frac{\beta}{\alpha F(A)}\right)\sum_{y \in A} \exp\left(\frac{\alpha v(y)}{\alpha F(A)}\right)} \\ &= \frac{\left(\exp\left(\frac{\alpha v(x) + \beta}{\alpha F(A)}\right)\right)}{\sum_{y \in A} \exp\left(\frac{\alpha v(y) + \beta}{\alpha F(A)}\right)}. \end{aligned}$$

Therefore, the pair $(\alpha v + \beta, \alpha F)$ also represents ρ, as required.

For uniqueness properties, first suppose (v, F) and (v', F') both represent ρ. By Theorem 2, there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that

$$\begin{aligned} v &= \alpha v' + \beta, \\ F &= \alpha F'. \end{aligned}$$

Thus, we have for any $x, y \in A$,

$$\frac{v(x)}{F(A)} - \frac{v(y)}{F(A)} = \frac{\alpha v'(x) + \beta}{\alpha F'(A)} - \frac{\alpha v'(y) + \beta}{\alpha F'(A)} = \frac{v'(x)}{F'(A)} - \frac{v'(y)}{F'(A)},$$

as desired. Next, suppose (v, F) represents ρ and there exists (v', F') such that

$$\frac{v(x)}{F(A)} - \frac{v(y)}{F(A)} = \frac{v'(x)}{F'(A)} - \frac{v'(y)}{F'(A)}.$$

This says that for any set A, there is a $\gamma_A \in \mathbb{R}$ such that

$$\frac{v'(x)}{F'(A)} = \frac{v(x)}{F(A)} + \gamma_A.$$

It follows that

$$\begin{aligned}
\frac{\exp\left(\frac{v'(x)}{F'(A)}\right)}{\sum_{y \in A} \exp\left(\frac{v'(y)}{F'(A)}\right)} &= \frac{\exp\left(\frac{v(x)}{F(A)} + \gamma_A\right)}{\sum_{y \in A} \exp\left(\frac{v(y)}{F(A)} + \gamma_A\right)} = \frac{\exp(\gamma_A) \exp\left(\frac{v(x)}{F(A)}\right)}{\exp(\gamma_A) \sum_{y \in A} \exp\left(\frac{v(y)}{F(A)}\right)} \\
&= \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\sum_{y \in A} \exp\left(\frac{v(y)}{F(A)}\right)} = \rho(a, A).
\end{aligned}$$

This establishes that (v', F') represents ρ.

D Proof of Theorem 3

First suppose that ρ is a relaxed IIA rule. From Equation (2) we see that for any $x, y \in A$, $\rho(x, A) \geq \rho(y, A)$ holds if and only if $v(x) \geq v(y)$. Since v does not depend on A, this proves Axiom 1. To prove Axiom 2, observe that for $x, y \in A$,

$$\ln \frac{\rho(x, A)}{\rho(y, A)} = \frac{v(x) - v(y)}{F(A)}.$$

From which it follows that if $x, y, x', y' \in A \cap B$ we have

$$\ln \frac{\rho(x, A)}{\rho(y, A)} \ln \frac{\rho(x', B)}{\rho(y', B)} = \frac{(v(x') - v(y'))(v(x) - v(y))}{F(A) F(B)}$$

and similarly

$$\ln \frac{\rho(x, B)}{\rho(y, B)} \ln \frac{\rho(x', A)}{\rho(y', A)} = \frac{(v(x') - v(y'))(v(x) - v(y))}{F(A) F(B)},$$

which proves Axiom 2. This establishes the necessity of the axioms.

Now suppose ρ is a random choice rule that obeys Axioms 1 and 2. We will define $v : X \to (0, \infty)$ and $F : \mathcal{A} \to (0, \infty)$ so that Equation (2) holds, from which it follows that ρ is a relaxed IIA rule. We first define

$$\alpha := 1 - \min_{z \in X} \ln \rho(z, A).$$

Since $\rho(z, A) \in (0, 1)$ whenever $z \in A$, it follows that $\alpha \in (1, \infty)$. Now define v by

$$v(x) = \alpha + \ln \rho(x, X).$$

Using $\alpha \in (1, \infty)$, we see that $v(x) \in (0, \infty)$ for all $x \in X$. It remains to define F. Start with $A \in \mathcal{A}$ for which there exist $x', y' \in A$ with $\rho(x', A) \neq \rho(y', A)$. By Axiom 1, we have $\rho(x', X) \neq \rho(y', X)$, so that $v(x') \neq v(y')$. Now define F by

$$F(A) = \ln\left(\frac{\rho(x', X)}{\rho(y', X)}\right)\left(\ln\left(\frac{\rho(x', A)}{\rho(y', A)}\right)\right)^{-1}.$$

We see that $F(X) = 1$. Also, by Axiom 2, the value of $F(A)$ is invariant to which x', y' are chosen, provided $\rho(x', A) \neq \rho(y', A)$ holds. Hence, for any $x, y \in A$ with $\rho(x, A) \neq \rho(y, A)$, we have

$$\ln \frac{\rho(x, A)}{\rho(y, A)} = \frac{1}{F(A)} \ln \frac{\rho(x, X)}{\rho(y, X)} = \frac{(v(x) - \alpha) - (v(y) - \alpha)}{F(A)} = \frac{v(x) - v(y)}{F(A)}.$$

This equation also holds whenever $\rho(x, A) = \rho(y, A)$, since both sides are then 1. Exponentiating and summing over all $x \in A$ we get

$$\sum_{x \in A} \frac{\rho(x, A)}{\rho(y, A)} = \sum_{x \in A} \exp\left(\frac{v(x) - v(y)}{F(A)}\right),$$

so that

$$\frac{1}{\rho(y, A)} = \frac{\sum_{x \in A} \exp\left(\frac{v(x)}{F(A)}\right)}{\exp\left(\frac{v(y)}{F(A)}\right)},$$

from which, inverting both sides, we get Equation (2) for any A on which ρ is non-uniform. For a set A on which ρ is uniform, Axiom 1 tells us that v is constant on A. In this case, Equation (2) holds for any non-zero value of $F(A)$, and we can simply set $F(A) = 1$.

E Examples

We now provide various examples promised in the text. Example 4 gives a class of an inadmissible family of functions. Example 5 demonstrates that Axiom 2 does not imply Axiom 1. And Example 6 shows that the power form for the family $\{G_A\}_{A \in \mathcal{A}}$ is not the only kind of relaxed IIA rule.

Example 6 may seem at first to contradict the proof of Theorem 1, which relied on using the power form for any admissible family of functions. If $B \in \mathcal{A}$ has at least three alternatives, then the proof showed that $G_B(\cdot)$ must be exactly the power form. However, if $B \in \mathcal{A}$ has two alternatives, the proof only established the weaker statement that, if ρ obeys the relaxed IIA equality, Equation (1), under admissible family $\{G_A\}_{A\in\mathcal{A}}$, then one could change $G_B(\cdot)$ to be a power form while maintaining admissibility and Equation (1) relative to ρ.[10] This weaker statement is sufficient to prove Theorem 1, but does not rule out the possibility of admissible families that do not use the power form for sets with two alternatives. Example 6 fills this gap by demonstrating that such admissible families do in fact exist.

Example 4. Let $B, C \in \mathcal{A}$ with $|B \cap C| \geq 3$. Suppose $\{G_A\}_{A\in\mathcal{A}}$ is a family of strictly increasing functions mapping $(0,\infty)$ to $(0,\infty)$. Further suppose that $G_B(t) = \alpha t$ and $G_C(t) = \beta t$ with $\alpha \neq \beta$. In order for $\{G_A\}_{A\in\mathcal{A}}$ to be admissible there must be at least one random choice rule ρ that satisfies Equation (1) using $\{G_A\}_{A\in\mathcal{A}}$. Let ρ be such a random choice rule. Let x, y, z be three distinct elements of $B \cap C$. Then for any $w, w' \in \{x, y, z\}$ we have

$$\alpha \frac{\rho(w,B)}{\rho(w',B)} = \beta \frac{\rho(w,C)}{\rho(w',C)}.$$

Since $\alpha \neq \beta$, and ρ is full support, it is immediate that neither α nor β can be zero. Therefore for any $w, w' \in \{x, y, z\}$

$$\frac{\rho(w,B)}{\rho(w',B)} = \frac{\beta}{\alpha} \frac{\rho(w,C)}{\rho(w',C)}. \tag{9}$$

Applying the above equation we can derive that

$$\frac{\rho(x,B)}{\rho(z,B)} = \frac{\rho(x,B)}{\rho(y,B)} \frac{\rho(y,B)}{\rho(z,B)} = \left(\frac{\beta}{\alpha}\right) \frac{\rho(x,C)}{\rho(y,C)} \left(\frac{\beta}{\alpha}\right) \frac{\rho(y,C)}{\rho(z,C)} = \left(\frac{\beta}{\alpha}\right)^2 \frac{\rho(x,C)}{\rho(z,C)}.$$

And since $\beta \neq \alpha$, that contradicts Equation (9) when $w = x$ and $w' = z$. Therefore, we have shown there exists no ρ that satisfies Equation (1) using $\{G_A\}_{A\in\mathcal{A}}$, and hence $\{G_A\}_{A\in\mathcal{A}}$ is inadmissible.

Example 5. Let $X = \{x, y, z\}$ and consider a random choice rule ρ defined as follows:

$$\begin{aligned}
\rho(x,\{x,y,z\}) &= 0.5, \rho(y,\{x,y,z\}) = 0.3, \rho(z,\{x,y,z\}) = 0.2, \\
\rho(x,\{x,y\}) &= 0.4, \rho(y,\{x,y\}) = 0.6, \\
\rho(y,\{y,z\}) &= 0.6, \rho(z,\{y,z\}) = 0.4, \\
\rho(x,\{x,z\}) &= 0.6, \rho(z,\{x,z\}) = 0.4.
\end{aligned}$$

[10] This nuance is glossed over in the proof sketch in Section 3, but is treated in detail in the full proof in the Appendix.

ρ does not obey Axiom 1 since $\rho(x,X) > \rho(y,X)$ and $\rho(x,\{x,y\}) < \rho(y,\{x,y\})$. We now want to show that ρ obeys Axiom 2. Let $A, B \in \mathcal{A}$ and suppose that $w, z, w', z' \in A \cap B$. We want to show that

$$\ln \frac{\rho(w,A)}{\rho(z,A)} \ln \frac{\rho(w',B)}{\rho(z',B)} = \ln \frac{\rho(w',A)}{\rho(z',A)} \ln \frac{\rho(w,B)}{\rho(z,B)} \tag{10}$$

If $A = B$, then Equation (10) holds immediately. If $w' = z'$ or $w = z$ the above equation again holds since both sides would be zero. From here on suppose $A \neq B$, $w' \neq z'$ and $w \neq z$. These conditions imply that either A or B contains exactly two alternatives. Since all four of w, z, w', z' must all be contained in a set with two alternatives, either (i) $w = w'$, $z = z'$ or (ii) $w = z'$, $z = w'$. In case (i) it is again immediately clear that Equation (10) holds. In case (ii) we have that

$$\begin{aligned} \ln \frac{\rho(w,A)}{\rho(z,A)} \ln \frac{\rho(w',B)}{\rho(z',B)} &= \ln \frac{\rho(w,A)}{\rho(z,A)} \ln \frac{\rho(z,B)}{\rho(w,B)} \\ &= \left(-\ln \frac{\rho(z,A)}{\rho(w,A)}\right)\left(-\ln \frac{\rho(w,B)}{\rho(z,B)}\right) \\ &= \left(\ln \frac{\rho(w',A)}{\rho(z',A)}\right)\left(\ln \frac{\rho(w,B)}{\rho(z,B)}\right), \end{aligned}$$

which establishes Equation (10).

Example 6. Let $X = \{x, y, z\}$. Define $G_X : (0,\infty) \to (0,\infty)$ by

$$G_X(t) = \begin{cases} 2t - 1 & \text{if } t \geq 1, \\ \frac{t}{2-t} & \text{if } t < 1, \end{cases}$$

and define $G_A(t) = t$ for all $A \neq X$. It is clear that $\{G_A\}_{A\in\mathcal{A}}$ is not a power form. So, we have only to check that $\{G_A\}_{A\in\mathcal{A}}$ is admissible. Fix any $\lambda \in \Delta X$. We now construct a random choice rule ρ which satisfies Equation (1) under $\{G_A\}_{A\in\mathcal{A}}$ and which agrees with λ on X. Set $\rho(x',X) = \lambda(x')$ for all $x' \in X$. Choose any two-element set $A = \{y', z'\}$ and without loss of generality suppose $\lambda(y') > \lambda(z')$. Let

$$\rho(y',A) = 1 - \frac{1}{2} \cdot \frac{\lambda(z')}{\lambda(y')} \text{ and } \rho(z',A) = \frac{1}{2} \cdot \frac{\lambda(z')}{\lambda(y')}.$$

It can be checked that ρ defined this way satisfies Equation (1) under $\{G_A\}_{A\in\mathcal{A}}$. Moreover, by varying $\lambda(y')$ and $\lambda(z')$, all possible (non-zero) probabilities can be generated on A. This means that we could have similarly constructed ρ by starting with any two-element choice set B and arbitrary probabilities on B. This establishes that $\{G_A\}_{A\in\mathcal{A}}$ is admissible.

F Proof of Proposition 1

We will need a preliminary result.

Lemma 2. *Fix some $A \in \mathcal{A}$, $\lambda \in \Delta A$, and $F : \mathcal{A} \to (0, \infty)$. For any random choice rule ρ with $\rho(\cdot, A) = \lambda$, the following are equivalent:*

1. *For all $B, C \in \mathcal{A}$,*

$$\left(\frac{\rho(x,B)}{\rho(y,B)}\right)^{F(B)} = \left(\frac{\rho(x,C)}{\rho(y,C)}\right)^{F(C)} \text{ for all } x, y \in B \cap C.$$

2. *There exists $w : X \backslash A \to (-\infty, +\infty)$ such that for any $B \in \mathcal{A}$ and $x \in A \cap B$,*

$$\rho(x,B) = \frac{\lambda(x)^{F(A)/F(B)}}{\sum_{y \in A \cap B} \lambda(y)^{F(A)/F(B)} + \sum_{y \in B \backslash A} \exp\left(\frac{w(y)}{F(B)}\right)},$$

and for any $B \in \mathcal{A}$ and $x \in B \backslash A$,

$$\rho(x,B) = \frac{\exp\left(\frac{w(x)}{F(B)}\right)}{\sum_{y \in A \cap B} \lambda(y)^{F(A)/F(B)} + \sum_{y \in B \backslash A} \exp\left(\frac{w(y)}{F(B)}\right)}.$$

Note that $0 < \rho(x,B) \leq 1$ and $\sum_{x \in B} \rho(x,B) = 1$ for all $B \in \mathcal{A}$, so that ρ defined in Part 2. of the lemma is a well-defined random choice rule.

Proof. We will first show $2 \Rightarrow 1$. Extend w to all of X by defining, for every $x \in A$,

$$w(x) = F(A) \ln \lambda(x).$$

This implies that for any $B \in \mathcal{A}$ and $x \in B$,

$$\rho(x,B) = \frac{\exp\left(\frac{w(x)}{F(B)}\right)}{\sum_{y \in B} \exp\left(\frac{w(y)}{F(B)}\right)}.$$

Therefore, for all $x, y \in B$ we get

$$\left(\frac{\rho(x,B)}{\rho(y,B)}\right)^{F(B)} = \exp(w(x) - w(y)),$$

and since the right-hand side of this equation does not depend on B, we get the equation in Part 1.

We next show that $1 \Rightarrow 2$. Fix any $z \in A$ and define $w : X \backslash A \to (-\infty, +\infty)$ by setting, for each $x \in X \backslash A$,

$$w(x) = F(X) \ln \left(\frac{\rho(x, X)}{\rho(z, X)} \times \rho(z, A)^{\frac{F(A)}{F(X)}} \right).$$

By construction we have $w(x) \in (-\infty, +\infty)$ for all $x \in X \backslash A$. Now choose any $B \in \mathcal{A}$ and $x \in B$. We know that for all $y \in B$,

$$\frac{\rho(y, B)}{\rho(x, B)} = \frac{\rho(y, X)}{\rho(x, X)}^{\frac{F(X)}{F(B)}},$$

which implies

$$\sum_{y \in B} \frac{\rho(y, B)}{\rho(x, B)} = \sum_{y \in B} \frac{\rho(y, X)}{\rho(x, X)}^{\frac{F(X)}{F(B)}}.$$

From this we can write

$$\rho(x, B) = \frac{\rho(x, X)^{\frac{F(X)}{F(B)}}}{\sum_{y \in B} (\rho(y, X))^{\frac{F(X)}{F(B)}}} = \frac{\rho(x, X)^{\frac{F(X)}{F(B)}}}{\sum_{y \in A \cap B} (\rho(y, X))^{\frac{F(X)}{F(B)}} + \sum_{y \in B \backslash A} (\rho(y, X))^{\frac{F(X)}{F(B)}}}.$$

The ratio equation in Part 1. implies that for all $y \in A$,

$$\rho(y, X) = \rho(z, X) \left(\frac{\rho(y, A)}{\rho(z, A)} \right)^{\frac{F(A)}{F(X)}}.$$

From this and the definition of w we get

$$\rho(x, B) = \frac{\rho(x, X)^{\frac{F(X)}{F(B)}}}{\sum_{y \in A \cap B} \rho(y, A)^{\frac{F(A)}{F(B)}} \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}} + \sum_{y \in B \backslash A} \exp\left(\frac{w(y)}{F(B)}\right) \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}}}.$$

If $x \in A$, we can rewrite this equation as

$$\rho(x, B) = \frac{\rho(x, A)^{\frac{F(A)}{F(B)}} \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}}}{\sum_{y \in A \cap B} \rho(y, A)^{\frac{F(A)}{F(B)}} \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}} + \sum_{y \in B \backslash A} \exp\left(\frac{w(y)}{F(B)}\right) \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}}}$$

$$= \frac{\rho(x,A)^{\frac{F(A)}{F(B)}}}{\sum_{y\in A\cap B}\rho(y,A)^{\frac{F(A)}{F(B)}} + \sum_{y\in B\setminus A}\exp\left(\frac{w(y)}{F(B)}\right)}.$$

If $x \notin A$, we can write

$$\begin{aligned}\rho(x,B) &= \frac{\exp\left(\frac{w(x)}{F(B)}\right) \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}}}{\sum_{y\in A\cap B}\rho(y,A)^{\frac{F(A)}{F(B)}} \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}} + \sum_{y\in B\setminus A}\exp\left(\frac{w(y)}{F(B)}\right) \times \frac{\rho(z,X)^{\frac{F(X)}{F(B)}}}{\rho(z,A)^{\frac{F(A)}{F(B)}}}} \\ &= \frac{\exp\left(\frac{w(x)}{F(B)}\right)}{\sum_{y\in A\cap B}\rho(y,A)^{\frac{F(A)}{F(B)}} + \sum_{y\in B\setminus A}\exp\left(\frac{w(y)}{F(B)}\right)},\end{aligned}$$

as desired. □

We can now complete the proof of Proposition 1.1. Fix $A \in \mathcal{A}$, $\lambda \in \Delta A$, and $x \in A, y \notin A$. First suppose that $\lambda(z) = \lambda(z')$ for all $z, z' \in A$. Let ρ be a relaxed IIA rule with $\rho(\cdot, A) = \lambda$. We want to show that $\rho(x,A) \geq \rho(x, A\cup\{y\})$. By Axiom 1, we know that $\rho(z, A\cup\{y\}) = \rho(x, A\cup\{y\})$ for any $z \in A$, from which,

$$1 = \sum_{z\in A\cup\{y\}} \rho(z, A\cup\{y\}) \geq |A|\rho(x, A\cup\{y\}),$$

and therefore

$$\rho(x, A\cup\{y\}) \leq \frac{1}{|A|} = \rho(x,A),$$

as desired.

Now suppose there exist $z, z' \in A$ with $\lambda(z) \neq \lambda(z')$. Fix any $F : \mathcal{A} \to (0,\infty)$, with $F(A) = 1$, and any $w_y \in (-\infty, +\infty)$. By Lemma 2, we know there exists a relaxed IIA rule ρ with $\rho(\cdot, A) = \lambda$ and

$$\rho(x, A\cup\{y\}) = \frac{\rho(x,A)^{1/F(A\cup\{y\})}}{\sum_{z\in A}\rho(z,A)^{1/F(A\cup\{y\})} + \exp\left(\frac{w_y}{F(A\cup\{y\})}\right)}. \quad (11)$$

Suppose $\rho(x,A) < \frac{1}{|A|}$. Then we can take limits to get

$$\lim_{F(A\cup\{y\})\to+\infty}\lim_{w_y\to-\infty}\rho(x, A\cup\{y\}) = \frac{\rho(x,A)^{1/F(A\cup\{y\})}}{\sum_{z\in A}\rho(z,A)^{1/F(A\cup\{y\})}} = \frac{1}{|A|}.$$

Thus, we can choose $F(A \cup \{y\})$ and w_y so that $\rho(x, A \cup \{y\}) > \rho(x, A)$, as desired.

Next suppose that $\rho(x, A) \geq \frac{1}{|A|}$. We first show that if this is true, then there exists $\tilde{F} \in (0, \infty)$ with

$$\frac{\rho(x, A)^{1/\tilde{F}}}{\sum_{z \in A} \rho(z, A)^{1/\tilde{F}}} > \rho(x, A). \tag{12}$$

To see why this holds, note that if $\tilde{F} = 1$, then

$$\frac{\rho(x, A)^{1/\tilde{F}}}{\sum_{z \in A} \rho(z, A)^{1/\tilde{F}}} = \rho(x, A).$$

Since the left-hand side of this equality is continuously differentiable in $\tilde{F}$, it suffices to prove that the derivative of the left-hand side with respect to $\tilde{F}$ is non-zero at $\tilde{F} = 1$. This derivative is readily calculated to be

$$-\frac{\rho(x, A)^{1/\tilde{F}}}{\tilde{F}^2} \left(\sum_{z \in A} \rho(z, A)^{1/\tilde{F}} \right)^2 \left(\sum_{z \in A} \rho(z, A)^{1/\tilde{F}} \left(\ln \rho(x, A) - \ln \rho(z, A) \right) \right),$$

which, evaluated at $\tilde{F} = 1$, is

$$-\rho(x, A) \left(\sum_{z \in A} \rho(z, A) \right)^2 \left(\sum_{z \in A} \rho(z, A) \left(\ln \rho(x, A) - \ln \rho(z, A) \right) \right).$$

The sign of this expression is the same as the sign of

$$\sum_{z \in A} \rho(z, A) \ln \rho(z, A) - \ln \rho(x, A).$$

Thus, the derivative is zero exactly when

$$\sum_{z \in A} \rho(z, A) \ln \rho(z, A) = \ln \rho(x, A) \geq \ln \frac{1}{|A|},$$

where the inequality uses our assumption that $\rho(x, A) \geq \frac{1}{|A|}$. But $\sum_{z \in A} \rho(z, A) \ln \rho(z, A)$ is the negative of the entropy of ρ on set A, and $\ln |A|$ is the maximum possible entropy of ρ on set A. The maximum entropy is obtained only when $\rho(z, A) = \frac{1}{|A|}$ for all $z \in A$, which contradicts our assumption of non-uniformity of ρ on A. This establishes Equation (7).

We can now finish the case $\rho(x, A) \geq \frac{1}{|A|}$. Setting $F(A \cup \{y\}) = \tilde{F}$ in Equation (6) gives

$$\rho(x, A \cup \{y\}) = \frac{\rho(x, A)^{1/\tilde{F}}}{\sum_{z \in A} \rho(z, A)^{1/\tilde{F}} + \exp\left(\frac{w_y}{\tilde{F}}\right)}.$$

Using Equation (7), we see that, by choosing w_y sufficiently negative, we can then arrange $\rho(x, A \cup \{y\}) > \rho(x, A)$, as desired.

We turn, finally, to the proof of Proposition 1.2. Fix any $A \in \mathcal{A}$, $\lambda \in \Delta A$, and distinct elements $x, y \in A$. Choose $F : \mathcal{A} \to (0, \infty)$ with $F(A) = 1$. From Lemma 2, we know there is a relaxed IIA rule ρ with $\rho(\cdot, A) = \lambda$ and

$$\rho(x, A \backslash \{y\}) = \frac{\rho(x, A)^{1/F(A \backslash \{y\})}}{\sum_{z \in A \backslash \{y\}} \rho(z, A)^{1/F(A \backslash \{y\})}}.$$

Suppose $\lambda(x) < \max_{z \in A \backslash \{y\}} \lambda(z)$, that is, there exists $z \in A \backslash \{y\}$ with $\rho(z, A) > \rho(x, A)$. If we take $F(A \backslash \{y\}) \to 0$, we get $\rho(x, A \backslash \{y\}) \to 0$. But we know that $\rho(x, A) = \lambda(x) > 0$, so there must be a sufficiently small value for $F(A \backslash \{y\}) > 0$ for which $\rho(x, A) > \rho(x, A \backslash \{y\})$, as desired. Next suppose $\lambda(x) > \frac{1}{|A|-1}$. Taking $F(A \backslash \{y\}) \to +\infty$, we get $\rho(x, A \backslash \{y\}) \to \frac{1}{|A|-1}$. Thus, there must be a sufficiently large value for $F(A \cup \{y\})$ for which $\rho(x, A \backslash \{y\}) < \rho(x, A)$, as desired.

We now prove the reverse direction. Suppose $\lambda(x) \geq \max_{z \in A \backslash \{y\}} \lambda(z)$ and $\lambda(x) \leq \frac{1}{|A|-1}$. This gives $\rho(x, A) \geq \rho(z, A)$ for all $z \in A \backslash \{y\}$, from which, by Axiom 1,

$$\rho(x, A \backslash \{y\}) \geq \rho(z, A \backslash \{y\})$$

for all $z \in A \backslash \{y\}$. Summing over z yields

$$(|A| - 1)\, \rho(x, A \backslash \{y\}) \geq \sum_{z \in A \backslash \{y\}} \rho(z, A \backslash \{y\}) = 1,$$

and therefore

$$\rho(x, A \backslash \{y\}) \geq \frac{1}{|A| - 1} \geq \rho(x, A),$$

as desired.

G Proof of Proposition 2

Let ρ be a relaxed IIA rule, then there must exist $v : X \to (0, \infty)$ and $F : \mathcal{A} \to (0, \infty)$ such that for all $A \in \mathcal{A}$ and $x \in A$, Equation (2) holds. Now let $A, B \in \mathcal{A}$ with $F(A) \geq F(B)$. We will show A is more stochastic than B. Choose any $x, y \in A \cap B$ such that $\rho(x, A) \geq \rho(y, A)$.

It is clear from Equation (2) that $\rho(x, A) \geq \rho(y, A)$ implies $v(x) \geq v(y)$. Again employing Equation (2),

$$\frac{\rho(x, A)}{\rho(y, A)} = \frac{\exp\left(\frac{v(x)}{F(A)}\right)}{\exp\left(\frac{v(y)}{F(A)}\right)} = \exp\left(\frac{v(x) - v(y)}{F(A)}\right)$$

and

$$\frac{\rho(x, B)}{\rho(y, B)} = \frac{\exp\left(\frac{v(x)}{F(B)}\right)}{\exp\left(\frac{v(y)}{F(B)}\right)} = \exp\left(\frac{v(x) - v(y)}{F(B)}\right).$$

Since $F(A) \geq F(B)$ and $v(x) - v(y) > 0$ we know that

$$\frac{\rho(x, A)}{\rho(y, A)} \leq \frac{\rho(x, B)}{\rho(y, B)}.$$

This establishes that $F(A) \geq F(B)$ implies that A is more stochastic than B. The statement of Proposition 2 follows since for any $A, B \in \mathcal{A}$, either $F(A) \geq F(B)$ or $F(B) \geq F(A)$ must hold.

References

Aczél, J. (1966). *Lectures on Functional Equations And Their Applications*. New York: Academic Press.

Attneave, F. (1954). Some Informational Aspects of Visual Perception. *Psychological review 61*(3), 183–193.

Barlow, H. (1961). Possible Principles Underlying the Transformation of Sensory Messages. In W. A. Rosenblith (Ed.), *Sensory Communication*, pp. 217–234. Cambridge: M.I.T. Press.

Barraclough, D. J., M. L. Conroy, and D. Lee (2004). Prefrontal Cortex and Decision Making in a Mixed-Strategy Game. *Nature neuroscience 7*(4), 404–410.

Berkowitsch, N., B. Scheibehenne, and J. Rieskamp (2014). Rigorously Testing Multialternative Decision Field Theory Against Random Utility Models. *Journal of experimental psychology. General 143*(3), 1331–48.

Bernhardt, B. C. and T. Singer (2012). The Neural Basis of Empathy. *Annual Review of Neuroscience 35*(1), 1–23.

Best, P. J., A. M. White, and A. Minai (2001). Spatial Processing in the Brain: The Cctivity of Hippocampal Place Cells. *Ann. Rev. Neurosci. 24*(0), 459–486.

Block, H. D. and J. Marschak (1960). Random Orderings and Stochastic Theories of Responses. In I. Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pp. 97–132. Stanford: Stanford University Press.

Caplin, A. and M. Dean (2015). Revealed Preference, Rational Inattention, and Costly Information Acquisition. *American Economic Review 105*(7), 2183–2203.

Carandini, M. and D. Heeger (2012). Normalization as a Canonical Neural Computation. *Nature Reviews Neuroscience* (November), 1–12.

Churchland, A. K., R. Kiani, R. Chaudhuri, X. J. Wang, A. Pouget, and M. N. Shadlen (2011). Variance as a Signature of Neural Computations during Decision Making. *Neuron 69*(4), 818–831.

David, H. and H. Nagaraja (2003). *Order Statistics* (3rd ed.). New York: Wiley.

Delgado, M. R., L. E. Nystrom, C. Fissell, D. C. Noll, and J. a. Fiez (2000). Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum. *Journal of neurophysiology 84*(6), 3072–3077.

Dorris, M. C. and P. W. Glimcher (2004). Activity in Posterior Parietal Cortex is Correlated with the Relative Subjective Desirability of Action. *Neuron 44*(2), 365–378.

Elliott, R., K. J. Friston, and R. J. Dolan (2000). Dissociable Neural Responses in Human Reward Systems. *The Journal of neuroscience : the official journal of the Society for Neuroscience 20*(16), 6159–65.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015). Stochastic Choice and Revealed Perturbed Utility. *Econometrica 83*(6), 2371–2409.

Glimcher, P. (2014). Understanding the Hows and Whys of Decision-Making: From Expected Utility to Divisive Normalization. *Cold Spring Harbor Symposia on Quantitative Biology 79*, 169–176.

Glimcher, P. W. and A. A. Tymula (2016). Expected Subjective Value Theory (ESVT): A Representation of Decision under Risk and Certainty.

Gul, F., P. Natenzon, and W. Pesendorfer (2014). Random Choice as Behavioral Optimization. *Econometrica 82*(5), 1873–1912.

Heeger, D. J., E. P. Simoncelli, and J. A. Movshon (1996). Computational Models of Cortical Visual Processing. *Proc.Natl.Acad Sci.U.S.A 93*(2), 623–627.

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

Hubel, J. and T. Wiesel (2004). *Brain and Visual Perception: The Story of a 25-Year Collaboration.* USA: Oxford University Press.

Huber, J., J. W. Payne, and C. Puto (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research 9*(1), 90–98.

Hunt, L. T., R. J. Dolan, and T. E. J. Behrens (2014). Hierarchical Competitions Subserving Multi-Attribute Choice. *Nature Neuroscience 17*(October), 1–14.

Kable, J. W. and P. W. Glimcher (2007). The Neural Correlates of Subjective Value During Intertemporal Choice. *Nature neuroscience 10*(12), 1625–1633.

Kandel, E. R., J. H. Schwartz, and T. Jessel (2000). *Principles of Neural Science* (4 ed.). New York: McGraw-hill.

Knutson, B., C. M. Adams, G. W. Fong, and D. Hommer (2001). Anticipation of Increasing Monetary Reward Selectively Recruits Nucleus Accumbens. *J Neurosci 21*(16), RC159.

Landauer, R. (1961). Irreversibility and Heat Generation in the Computing Process. *IBM Journal of Research and Development 5*(July), 183–191.

Levy, D. J. and P. W. Glimcher (2012). The Root of All Value: A Neural Common Currency for Choice. *Current Opinion in Neurobiology 22*(6), 1027–1038.

Louie, K., L. E. Grattan, and P. W. Glimcher (2011). Reward Value-Based Gain Control: Divisive Normalization in Parietal Cortex. *Journal of Neuroscience 31*(29), 10627–10639.

Louie, K., M. W. Khaw, and P. W. Glimcher (2013). Normalization is a General Neural Mechanism for Context-Dependent Decision Making. *Proceedings of the National Academy of Sciences of the United States of America 110*(15), 6139–44.

Louie, K., T. LoFaro, R. Webb, and P. W. Glimcher (2014). Dynamic Divisive Normalization Predicts Time-Varying Value Coding in Decision-Related Circuits. *The Journal of Neuroscience 34*(48), 16046–16057.

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

Mandl, F. (1988). *Statistical Physics* (2nd ed.). John Wiley & Sons.

Marley, A. A. J., T. N. Flynn, and J. J. Louviere (2008). Probabilistic Models of Set-Dependent and Attribute-Level Best-Worst Choice. *Journal of Mathematical Psychology 52*(5), 281–296.

Matějka, F. and A. McKay (2014). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review 105*(1), 1–55.

Mattsson, L.-G. and J. W. Weibull (2002). Probabilistic Choice and Procedurally Bounded Rationality. *Games and Economic Behavior 41*(1), 61–78.

McFadden, D. (1978). Modelling the Choice of Residential Location. In S. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models*, Volume 673, pp. 75–96. Amsterdam: North-Holland.

Platt, M. L. and P. W. Glimcher (1999). Neural Correlates of Decision Variables in Parietal Cortex. *Nature 400*(6741), 233–238.

Rieskamp, J., J. 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.

Schwartz, O. and E. P. Simoncelli (2001). Natural Signal Statistics and Sensory Gain Control. *Nature neuroscience 4*(8), 819–825.

Shadlen, M. N. and W. T. Newsome (1994). Noise, Neural Codes and Cortical Organization. *Current Opinion in Neurobiology 4*(4), 569–579.

Shannon, C. E. (1948). A Mathematial Theory of Communication. *Bell System Technical Journal 27*(3), 379–423.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics 69*(1), 99–118.

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research 16*(2), 158–174.

Sims, C. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy 49*, 317–356.

Sims, C. A. (2003). Implications of Rational Inattention. *Journal of Monetary Economics 50*(3), 665–690.

Swait, J. and A. A. J. Marley (2013). Probabilistic Choice (Models) as a Result of Balancing Multiple Goals. *Journal of Mathematical Psychology 57*(1-2), 1–14.

Thurstone, L. (1927). A Law of Comparative Judgment. *Psychological Review 34*(4), 273–286.

Webb, R., P. W. Glimcher, and K. Louie (2014). Rationalizing Context-Dependent Preferences: Divisive Normalization and Neurobiological Constraints on Choice. *Working Paper*, 1–56.