ANNUAL REVIEWS

Normative Principles for Decision-Making in Natural Environments

Christopher Summerfield and Paula Parpart

Department of Experimental Psychology, University of Oxford, Oxford OX2 6GG,
United Kingdom; email: christopher.summerfield@psy.ox.ac.uk

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Abstract

The decisions we make are shaped by a lifetime of learning. Past experience
guides the way that we encode information in neural systems for percep-
tion and valuation, and determines the information we retrieve when mak-
ing decisions. Distinct literatures have discussed how lifelong learning and
local context shape decisions made about sensory signals, propositional in-
formation, or economic prospects. Here, we build bridges between these
literatures, arguing for common principles of adaptive rationality in per-
ception, cognition, and economic choice. We discuss how a single common
framework, based on normative principles of efficient coding and Bayesian
inference, can help us understand a myriad of human decision biases, in-
cluding sensory illusions, adaptive aftereffects, choice history biases, central
tendency effects, anchoring effects, contrast effects, framing effects, congru-
ency effects, reference-dependent valuation, nonlinear utility functions, and
discretization heuristics. We describe a simple computational framework for
explaining these phenomena.

Contents

The mind of man is far from the nature of a clear and equal glass, wherein the beams of things should reflect according to their true incidence; nay, it is rather like an enchanted glass, full of superstition and imposture.

—Francis Bacon, *Instauratio Magna*

1. INTRODUCTION

Faced with a complex world, how should we think and act? This timeless question animates disciplines as diverse as philosophy, biology, and statistics. Here, we consider this question in light of empirical research describing how human behavior is structured and implemented in the brain (in psychology and neuroscience) and theoretical research aiming to derive normative principles for behavior and build intelligent agents (in cognitive science, economics, and machine learning). A theme that unites these fields is the possibility to define whether an agent's behavior is rational or optimal, given a specified objective in the environment and constraints on the availability of information or computational resources. Specifying an optimal policy is useful in the behavioral sciences because it provides a benchmark against which empirical phenomena can be compared and understood. However, the theoretical question of how to define optimal behavior, and the empirical question of whether the behavior approaches a normative standard—or is instead biased and irrational—has fueled controversy for more than half a century (Gershman et al. 2015, Simon 1955).

A normative theory makes two explicit commitments. First, it assumes an objective that an agent is seeking to achieve. In machine learning, this objective is specified as the cost function

that the agent is seeking to minimize. However, behavioral scientists often make do with more informal assumptions about the goals of humans or other animals. For example, an ecologist might explain how a foraging animal allocates time among food patches by assuming that it is maximizing energetic resources (Stephens & Krebs 1986), an economist might understand patterns of risky investment by assuming that an agent wishes to maximize their expected utility (Prelec 1998), or a psychologist might understand the latencies of sensory judgments by assuming that observers wish to maximize their accuracy per unit time (Bogacz 2007). Intellectual gridlock over normative claims often arises because a behavior may be rational under one definition but irrational under another. In other words, it can be unclear whether deviations from a normative standard are the fault of the agent (for failing to pursue their objective effectively) or of the researcher (for misspecifying the objective in the first place). A related problem arises in machine learning because even a fully specified cost function is usually only a proxy for the behavior that the designer actually wants to elicit. Thus, we might successfully train an artificial intelligence system to solve a problem but find that it learns to exploit loopholes or acquires harmful solutions. This is called the problem of value alignment (Christian 2020).

The second commitment concerns the knowledge to which the agent has access and the constraints under which it is encoded. Trivially, we cannot accuse an agent of being irrational for failing to act upon signals that it never sensed, has not stored in memory, or cannot access. A normative theory thus has to account for the sensitivity of our perceptual systems and the fidelity of our memory systems. These factors can be shaped by the hard constraints of physiology and the body (e.g., one has more photoreceptors at the fovea than the periphery) but also depend critically on the experience history of the organism. The way we learn is shaped by our relative exposure to different information, allowing us to become more expert in some domains than in others. This in turn can determine how humans and other animals think and act in laboratory studies, because participants will inevitably bring past knowledge, beliefs, and competences to bear on a task. In order to understand behavior, and to clarify its normative status, we thus need to understand the context that shaped the organism under study.

In the next section, we discuss diverse approaches to understanding human rationality in the context of natural environments. Subsequently, we propose an integrated framework drawing upon these accounts, and show how it can explain a wide variety of psychological phenomena.

2. ADAPTIVE RATIONALITY IN PERCEPTION, COGNITION, AND ECONOMIC DECISIONS

The idea that rational behavior can only be interpreted in the light of natural experience has motivated diverse theoretical perspectives in the cognitive sciences, going by names such as natural systems analysis (Geisler & Ringach 2009), adaptive rationality or ecological rationality (Todd & Gigerenzer 2012), rational inattention (Sims 2003), bounded rationality (Simon 1955), and Bayesian rationality (Oaksford & Chater 2009). In each case, researchers have argued that apparent human failures of perception, reasoning, or reward maximization are adaptive, in that they maximize the objective of the agent after accounting for the statistics and structure of the world, the physical limits on neural information processing, and the nature of biological cost functions. Each of these perspectives has proved controversial. We begin by summarizing the debates that have arisen in each domain.

2.1. Natural Systems Analysis

Illusions of perception reveal that humans often misconstrue the world. **Figure 1a** depicts two railway tracks that appear to be at different angles, even though the image is repeated (related
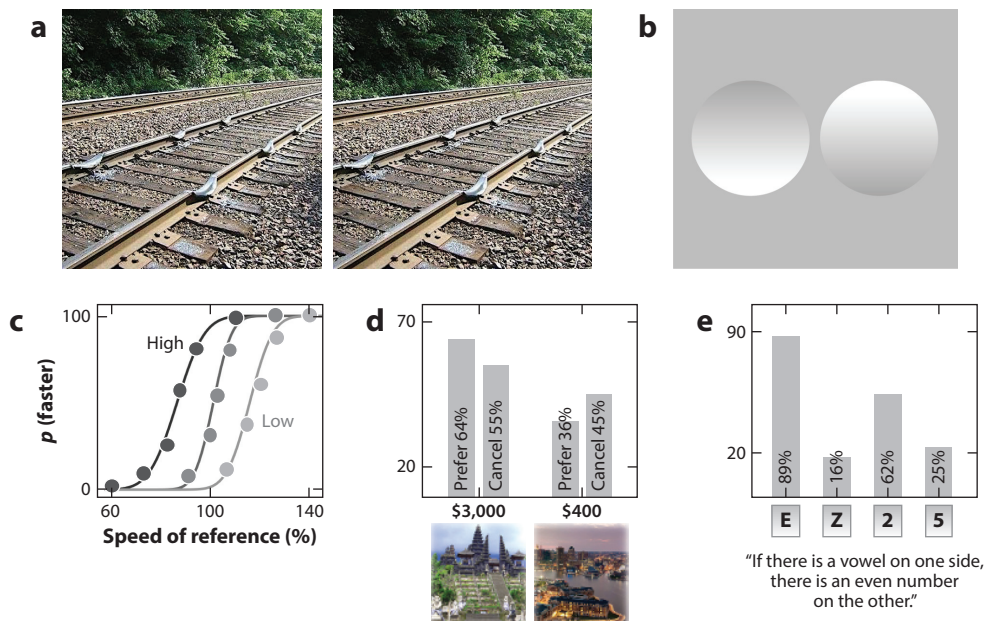
Figure 1

(*a*) An illusion of relative perception. The tracks on the right seem tilted clockwise in relation to those on the left, when in fact both images are the same. (*b*) The shape-from-shading illusion. The leftmost disc appears concave and the rightmost convex, but without additional assumptions the opposite interpretation is equally likely. (*c*) The probability of reporting that a probe grating was faster than a standard in high-, medium-, and low-contrast conditions (*dark*, *medium*, and *light gray lines*). (*d*) Participants heard about two offers, such as holidays, one of which had salient positive and negative features (exciting but expensive) whereas the other was more average on relevant dimensions. They decided to both accept and reject the first option. Percentages are from Shafir (1993, problem 1). (*e*) Percentages of participants who elected to turn over each card to verify the rule underneath. Data in panel *c* are schematic but follow Stone & Thompson (1992). Data in panel *e* are from the meta-analysis of the card selection task described by Oaksford & Chater (1994).

to the leaning tower illusion; Kingdom et al. 2007). **Figure 1*b*** displays the shape-from-shading illusion, which to most observers appears as a concave disc on the left and a convex disc on the right, even though the converse interpretation is equally plausible on the basis of sensory data alone (Ramachandran 1988). **Figure 1*c*** plots data showing that stimuli with lower contrast are typically perceived to move more slowly than those with higher contrast, even if their objective speeds are matched (Stone & Thompson 1992).

Natural systems analysis accounts for these perceptual phenomena as arising from the statistics of sensation (Geisler & Ringach 2009). These statistics are determined by the physical properties of the natural world, which is predictably populated by rivers, cities, trees, birds, and faces, all of which occur in scenes with stereotyped gestalt structure. Viewed from a typical human vantage point, natural scenes have well-characterized regularities, such as the tendency for light to come from above, for contrast to be lower at the top (far away) than at the bottom (close by) of a natural image, for more contours to be tilted cardinally than obliquely, or for a majority of objects to move slowly or not at all. A rational agent that exists in a world with these regularities should take them into account when making perceptual inferences (Geisler 2008).

One popular framework for explaining these perceptual anomalies appeals to Bayesian inference (Kersten et al. 2004). Bayesian models conceive of mental states as probability distributions

over beliefs and propose that inference occurs as if agents combine prior beliefs and sensory like-lihoods according to Bayes's rule. In the cases illustrated here, misperceptions may be rational for an agent that encodes these natural regularities as a prior belief. For example, perceiving the left disc in **Figure 1***b* as concave and the right as convex is optimal for an observer who combines the ambiguous sensory data with a prior belief that light comes from above. More subtle aspects of past experience may also shape the way we see the world. For example, the shape-from-shading illusion is strongest if the impression is given that stimuli are lit not from directly overhead but rather from the upper left. Most individuals position themselves so that their right hand does not cast shadow on an object it is holding; their prior is thus that the sun is to their left. Indeed, this left-side bias is stronger in right- than left-handed observers (Sun & Perona 1998). The contrast–speed effect discussed above can be explained if observers harbor a prior belief that most objects tend to move slowly or not at all. When estimating speed, the sensory evidence afforded under low contrast is weaker, so posterior estimates are more influenced by the prior (Stocker & Simoncelli 2006).

Thus, natural systems analysis makes the argument that sensory illusions, rather than revealing failures of our sensory systems, demonstrate that they are exquisitely tuned to deal with natural statistics (Barlow 1961, Geisler 2008, Simoncelli 2003). In other words, departures from normative behavior are a lens through which to examine the nature of the environment itself. However, this rhetorical turn conditions claims of rationality on hard-to-validate assumptions about the natural environment (do we really believe a priori that objects move slowly?). This difficulty of deciding whether we are studying the nature of behavior in a given environment, or the nature of the environment given (optimal) behavior, has provoked questions over the falsifiability of the approach (Bowers & Davis 2012, Jones & Love 2011). This class of argument is of course not new; the liberties afforded in specification of the prior were among the very earliest critiques leveled at Bayesian methods more than 200 years ago (McGrayne 2011).

2.2. Ecological Rationality

When humans make cognitive decisions about logical propositions, vignettes, or numerical data, our judgment often defies basic logic, ignores fundamental statistical precepts, and fails to max-imize test accuracy even in incentive-compatible settings. Instead, we tend to use heuristics or mental shortcuts when evaluating propositional information, to ignore sample sizes or base rates when evaluating probabilities, and to seek confirmatory evidence where falsification is required to draw certain conclusions (Kahneman 2012, Kahneman et al. 1982, Tversky & Kahneman 1974). Consider the example in **Figure 1***d*, where participants were offered the choice between two holi-day destinations which were approximately equally preferred because one was exotic but expensive (e.g., for US-based participants, Bali) and the other was less exciting but cheaper (e.g., Baltimore). When asked which they wanted to accept, on average participants chose Bali (presumably because it was exotic), but when asked which of the two options they wanted to reject, on average they also chose Bali (presumably because it was expensive). In other words, participants simultaneously preferred and dispreferred one of the options, a logical impossibility (Shafir 1993). Another classic demonstration is the card selection task shown in **Figure 1***e*, where participants were asked to turn over two cards to verify the truth of the statement that "if there is a vowel on one side, there is an even number on the other." Participants typically failed to seek key falsifying evidence (Wason 1960). Finally, participants often neglected part of the available information when making judg-ments, especially when combining multiple cues with potentially differing importance, resorting to stereotyped heuristics instead. Two well-known heuristics are those that treat all cues as equally important (equal weighting heuristics) and those that rely on a single cue (one-reason heuristics). For example, when asked to judge which of two cities is larger, participants may rely on a simple rule of thumb, such as whether it has a prominent football team (Gigerenzer & Gaissmaier 2011).

Once again, the point of contention here is not over the consistency or replicability of human bias but over its normative status. By definition, decision policies that are biased or deliberately underfitted to existing data (by discarding relevant information) are more likely to err—and thus, for a class of incentive-compatible judgement, less likely to maximize reward. However, in analogy to the natural systems analysis described above, proponents of ecological rationality argue that decisions are adapted to the structure of the world. Rather than revealing a fundamental irrationality in human cognition, these findings can be interpreted as evidence for rational behavior in a world in which information is costly to acquire (Oaksford & Chater 2009) and resources are finite (Todd & Gigerenzer 2012).

To illustrate these arguments, first consider the Wason card selection task. Participants fail to turn over cards that would falsify the contention, consistent with a large literature indicating that humans suffer from confirmatory decision biases (Nickerson 1998). However, the theory of optimal data selection (Oaksford & Chater 2003) argues that it is a property of the natural world (and presumably of the way we express ourselves) that the probability of any randomly chosen statement being true is lower than the probability of that statement being false. For example, out of all statements with the form "X is the current President of the United States," there is at most one X for which that statement is true and a theoretically infinite number of Xs for which it is false. Thus, the most efficient way to verify this statement would be to ask whether X is the current President, rather than systematically falsifying all other possible alternatives, which would be very time-consuming. Optimal data selection argues that this bias to sample data in the most resource-efficient way spills over into performance of logical tasks like that shown in **Figure 1e**, making participants more likely to rely on confirmatory strategies.

A different argument has been made to explain the reliance on heuristic decision strategies. Proponents of ecological rationality highlight evidence that heuristics can be superior when they match the structure of the environment. In an ever-changing and unpredictable world, our decision policies must be adapted not only to past sensory data (in machine learning terms, the training distribution) but also to a possible future (the test distribution). Policies that deliberately discard information (e.g., by relying on a single variable for making complex predictions, such as "take the best") might confer a robustness on decision policies that makes them well suited to the open-ended nature of real environments, because simpler policies often transfer better to novel settings (Volz & Gigerenzer 2012). As such, heuristics are regularized versions of optimal policies, conferring generalizability on our decisions and guarding against overfitting to extant experience (Parpart et al. 2018).

2.3. Rational Inattention

Finally, an enduring debate concerns human economic behavior—whether the decisions we make between consumer goods and monetary gambles are rational and reward maximizing or not. While the edifice of economic theory remains premised on a utility-maximizing *Homo economicus*, many regard this view as a convenient fiction, preferring to focus instead on the "misbehavior" of human economic choices (Lichtenstein & Slovic 2006, Thaler 2016). This view is supported by failures of utility maximization in choices among goods and gambles, idiosyncratic reversals of preference that depend on the framing of a risky choice, or that occur in the presence of irrelevant alternatives, or that disclose instability of preference through systematic intransitivity of choice.

We illustrate with a handful of phenomena drawn from a vast empirical literature, focusing on descriptive characterizations of human utility functions from monetary lotteries. The resulting data can be fitted with functions that distort objective monetary values or event probabilities during the mapping onto their subjective counterparts in ways that explain failures to make

choices that maximize expected value (Kahneman & Tversky 1979, Prelec 1998). For example, participants might choose between a sure bet of $3,000 and an 80% chance of $4,000 (or otherwise nothing). In these settings, participants often show a preference for safer lotteries with lower expected value, a phenomenon known as risk aversion. Risk aversion can be accounted for by a concave utility function, that is, one that assumes relatively smaller gains in utility for equivalent changes in larger sums (diminishing marginal utility). By assuming a specific form for utility and probability functions in the domain of both gains and losses, we can account for a wide range of otherwise idiosyncratic choices (Kahneman & Tversky 1979). For example, participants will typically only accept lotteries involving a 50:50 chance of gaining x_1 or losing x_2 when the amount x_1 is approximately twice x_2 (Tversky & Kahneman 1981). This entails that losses are subjectively weighted more heavily in choices, a phenomenon known as loss aversion. Loss aversion can lead to the same prospect being either preferred or dispreferred according to whether it is presented as a gain or a loss. For example, participants may prefer gamble A to B when presented as the chance of winning money, but when endowed with a larger sum and asked to choose between two potential losses that yield identical gambles, they prefer B over A.

These phenomena suggest that preferences are unstable and violate axioms of rational choice (von Neumann & Morgenstern 1944). Moreover, an agent whose preferences reverse systematically is susceptible to having money pumped from them by a Dutch book, or a set of prospects for which they are guaranteed to lose on average. Nevertheless, adaptively rational explanations have been offered for all of these phenomena and many others. For example, loss aversion may be rational for biological agents for whom resource scarcity is more harmful than resource abundance is beneficial, because of the ever-present possibility of death or the limited opportunities for caching (McDermott et al. 2008). It is well known that risk attitudes will change dynamically as proximity to a critical threshold approaches, such as calorific resources required during a cold night, in a way that may be adaptive for survival (Caraco 1981). Similar effects occur when humans gamble to reach a monetary threshold (Kolling et al. 2014).

Why might internal representations of probability and value be distorted despite evolutionary pressure for reward-maximizing decisions? One important intuition is that some phenomena may be rational for an agent whose cognitive systems have finite precision, because they have limited numbers of neurons (Dasgupta & Gershman 2021, Franconeri et al. 2013) or noisy inference (Drugowitsch et al. 2016). For example, the concave mapping of monetary value to utility that is a ubiquitous property of econometric functions is adaptive if lower-valued goods are more frequently encountered than higher-valued goods (as seems plausible), even in the absence of any assumptions about risk preference (Khaw et al. 2018). If valuation is conceived as a sampling process, and preferences are dynamically constructed by retrieving memories of past experiences to a current prospect, then utility and probability functions will be naturally tailored to match previously encountered distributions of value and probability. These distributions may favor low values and extreme probabilities, thereby accounting for idiosyncrasies in value (Bhui & Gershman 2018, Heng et al. 2020, Stewart et al. 2006). A different view suggests that biases at the extremities of the probability weighting function—whereby high and low probabilities are biased away from certainty—are a rational strategy for a noisy decision maker, because it corrects for an estimation bias that occurs when variability in the inference process is irreducible (Steiner & Stewart 2016).

3. AN INTEGRATIVE FRAMEWORK FOR ADAPTIVE RATIONALITY

Section 2 has summarized instances of irrational or suboptimal phenomena—erroneous perception of sensory stimuli, fallacies of logical reasoning, or failures to maximize expected utility—selected from across sensory neuroscience, cognitive psychology, and behavioral economics. In

each case, it has been argued that decisions may be adaptively rational under a slightly different specification of the problem agents evolved to solve. This point is related to the no free lunch theorem, which states that any two optimization algorithms are equivalent when their performance is averaged across all possible problems (Wolpert 1996).

In this section, we focus on the nature of the interrelationship between the purportedly irrational behaviors and the rational explanations that they have attracted in different fields. Casting an eye over the theoretical frameworks for which the principles of adaptive rationality have been invoked, we show that they differ markedly in functional form—there are Bayesian models in which the objects of computation are belief distributions; there are econometric models that posit nonlinear utility functions or invoke sampling from long-term memory; and there is an appeal to the deployment of heuristics, whose implementation is left largely open by their advocates. Can we identify equivalences among the solution concepts invoked in different domains—signals, tuning, likelihoods, reasons, utility? How should we think of the prior in a Bayesian model? Heuristics are often criticized for being an ad hoc bag of tricks—can we express them using computational tools, or hope to understand their neural origins? Where might a utility function reside in the brain, and how is it expressed? In what follows, we hope to shed light on these issues.

3.1. Differences Between Perception and Valuation

One challenge is that the processes underlying decisions about percepts and prospects feel very different. In perception, the quantities about which decisions are made (e.g., color or pitch) are usually some lawful function of available sensory input. For economic choices, the decision-relevant quantities (such as the calorific content of a food item or the worth of a consumer product) have to be learned by experience, by being encoded in a value function or retrieved from a model of the world (Daw et al. 2005). Correspondingly, the sources of noise or variability that corrupt perception and valuation may be different. One useful terminology distinguishes early (or Thurstonian) noise, which corrupts the encoding of sensory signals, from late (or Brunswickian) noise, which pertains to their mapping onto choices via a value function (Juslin & Olsson 1997). During psychophysical choice, errors arise primarily because of an inability to adequately detect, discriminate, or recognize a sensory stimulus, whereas the consequences of acting one way or another—conditional on the perceived signals—are typically transparent. During value-guided choice, the converse is usually true, in that the difficulty of choosing between supermarket products or between insurance policies is not because of an inability to sense the item (its packaging or the relevant numerical information) but rather because of uncertainty about its worth (Summerfield & Tsetsos 2012). In passing, we note that reinforcement learning models that are now popular computational theories in neuroscience assume that rewards and sensory observations are entirely different currencies processed by different input channels (Juechems & Summerfield 2019). Indeed, very different brain networks tend to be discussed in the study of perceptual and economic decisions irrespective of the species (Hanks & Summerfield 2017).

Nevertheless, we hope to offer a united account of adaptively rational decision-making that spans domains (perceptual, cognitive, and economic) and theoretical frameworks (Bayesian, econometric, and heuristic), bringing them together under a common set of principles that are characteristic of information processing in humans and other animals in the natural world. Ideally, such an account would be grounded in our understanding of cognitive systems mediating memory, attention, expectation, and motivation, but also expressed in a neurally plausible framework (invoking, for example, populations of neurons, their connectivity, and their dynamics). As a result, the framework will offer sufficient implementational detail and clarity for use by those who seek to build brains as well as understand them.

4. MODELING DECISIONS WITH NEURAL POPULATIONS

The brain is host to a dizzying array of cell types and transmitters arranged in stereotyped circuits across a patchwork of interconnected cortical and subcortical regions. In order to build models that are sufficiently constrained as to offer insights into behavior of the whole organism, we have to make simplifying assumptions. A popular idealization of brain function is to assume that information is encoded by a homogeneous class of processing units (or simulated neurons) and that its response to an input x can be summarized with a single number (or average firing rate). We can thus characterize the sensitivity of each neuron to the external world as a static tuning function, that is, a description of how it responds to features such as orientation or shape, or even to more abstract quantities such as the number of items in an array (Nieder & Dehaene 2009) or the reward probability associated with different pictures (Knudsen & Wallis 2021). How neurons encode information (and thus the shape of their tuning function) depends on their connectivity to sensory receptors or other neurons, which are often modeled as synaptic efficacies or connection weights. Depending on the research question or modeling goals, these weights may either be assigned values by hand or be trained by an optimization principle such as gradient descent (Simoncelli 2009).

To borrow a popular formalism, a stimulus x in the external world is encoded by a population of neurons with bell-shaped tuning curves such as those displayed in **Figure 2a**. We can consider
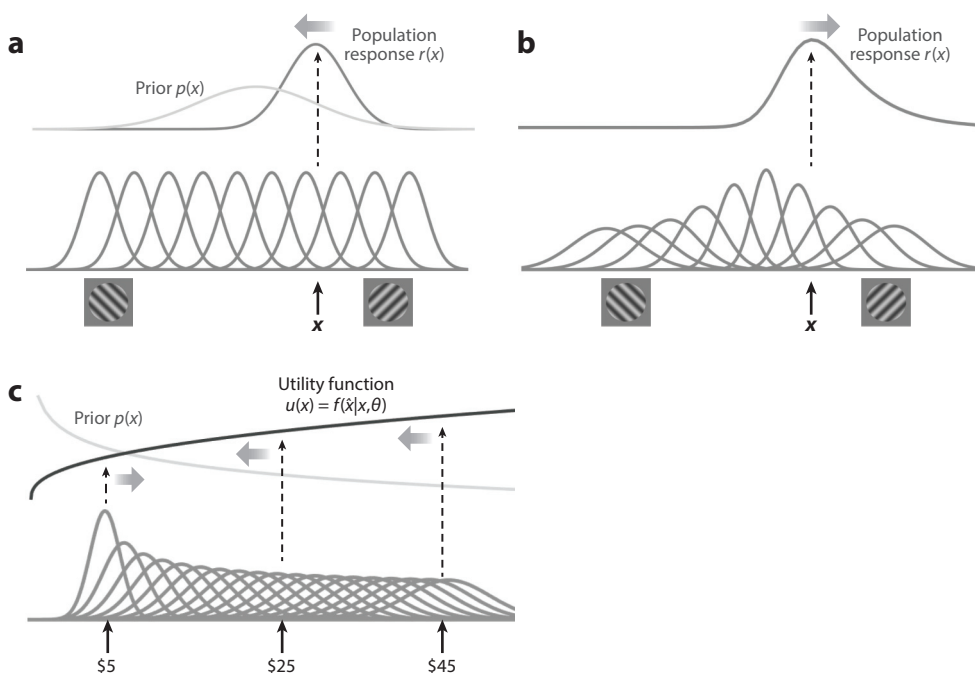
Figure 2

(*a*) Schematic of the population coding model. Neurons code for input features, such as a sensory property or monetary value, with a bell-shaped tuning function. A stimulus x elicits a population response $r(x)$. Decoded estimates are biased toward the mean of a prior distribution over likely features $p(x)$. (*b*) As in panel *a*, but with inhomogeneous tuning. The response distribution $r(x)$ is skewed away from the point at which tuning curves are sharpest, so that decoded estimates are biased away from this point. (*c*) These principles can be combined to explain nonlinear (concave) utility functions for economic goods or gambles.

the evoked activity of the neural population $r(x)$ to be a measurement of x. This neural response can be decoded into an output y, which might be a reconstructed estimate of the input itself (e.g., $\hat{x}$) or a probability associated with a categorical choice (such as pressing a left or right lever). We can thus describe the full model as implementing a function $y = f(x, \theta)$, where θ denotes parameters that control how information flows through the model—for example, setting the efficacies that determine the scale and width of the tuning functions. This framework thus allows us to model the process, assumed by both psychophysics and behavioral economics, in which a signal x is represented as a subjective or latent psychological quantity $\hat{x} = f(x)$, such as an internal perceptual or utility signal. Thus, the model allows us to formalize the likely source of distortions in these psychometric or econometric functions. The framework employed here is one of several that have been employed to study adaptive effects at the population level (for a review, see Weber et al. 2019).

4.1. Bayesian Inference and Attractive Choice Biases

Under simple assumptions, the model class described above can be used to implement Bayesian inference. We can conceive of the population activity $r(x)$ as encoding a likelihood distribution $p(r|x)$, which may be combined with a prior $p(x)$ to produce a posterior estimate $p(y|x)$ according to Bayes's rule. The prior distribution $p(x)$, which represents the experience history of the organism (e.g., frequencies of encounters with various sensory features, quantities, or values), can be envisaged to arrive from a separate memory system as a distinct input to the decoder. Note that the encoding model has the property that if either the input stimulus x itself is less certain or the tuning curves are wider, then the resulting likelihood function will be more dispersed. This means that, after combination with a prior distribution $p(x)$, the likelihood will hold less sway and so decisions will tend toward the mean of the prior when signals are more uncertain, the hallmark of Bayesian inference. Biases in perception can thus arise due to prior beliefs about the probability of occurrence of different features, such as a prior belief about the slowness of objects or the cardinality of contours, that were acquired in the natural world. This offers a natural explanation for effects observed in lab-based studies, such as a bias toward perceiving cardinal angles (Girshick et al. 2011) or slower speeds (Stocker & Simoncelli 2006). The same Bayesian explanation can apply to perceptual magnet effects, where perception is attracted toward prototypical category members. For example, observers are more sensitive to the distinction between blue and green (which crosses a category boundary) than they are to equivalent distinctions between shades of green or shades of blue; similar effects are observed for other stimuli, such as vowels or even faces. Such findings can be explained if the prior is bimodal with a peak at each prototype, assuming that blue and green (e.g., the grass and the sky) are more common than turquoise (Feldman et al. 2009).

Many prior beliefs may form through long-run experience with natural statistics. However, our expectations vary constantly with the local stimulation history (temporal context) and the backdrop provided by additional objects or items in a scene (spatial context). Accordingly, judgments about sensory features, numbers, and economic values tend to be attracted toward the mean of the recent past, a phenomenon called a contraction bias or central tendency effect (Hollingworth 1910, Petzschner et al. 2015). One instance of this phenomenon is illustrated in **Figure 2a**, where the prior peaks in the middle of feature space and estimates regress toward the mean. For example, when reproducing time intervals sampled from different ranges, estimates will regress toward the mean of the range, with stronger regression when estimates are more uncertain (Jazayeri & Shadlen 2010). Mice exhibit similar biases when discriminating the loudness of two successive auditory stimuli: During the delay period, perception of the first stimulus drifts toward the mean

of prior stimulation, biasing the comparison judgment (Akrami et al. 2018). Similar effects play out when participants make repeated decisions over multiple independent trials (such as judging the tilt of successive gratings). Another common observation is that successive judgments tend to be temporally congruent; that is, decisions are attracted toward the context implied by immediately past information. This occurs even where stimuli are sampled independently from trial to trial, rendering the past context irrelevant for current decisions (Abrahamyan et al. 2016, Akaishi et al. 2014, Fischer & Whitney 2014, Urai et al. 2019, Yu & Cohen 2009). These attractive choice history biases or sequential effects are at least partly perceptual in nature (Akaishi et al. 2014), although they may also be driven by a tendency to repeat the same response, an effect also known as the status quo bias (Fleming et al. 2010). Bayesian explanations for these phenomena appeal to the fact that in the natural world input signals tend to be nonstationary (drifting), so Bayesian agents with a prior belief that signals will be correlated in time will exhibit sequential biases (Yu & Cohen 2009).

Decisions are often made on the basis of sequential samples of evidence, for example, when adjacent frames hint at the likely direction of motion of a cloud of moving dots (Britten et al. 1992), or where successive cues indicate the probability that a given response will be rewarded (Basten et al. 2010, Yang & Shadlen 2007). A common finding in these tasks is that unequal weight is given to samples at either the beginning or the end of the sequence (Kiani et al. 2008, Lange et al. 2019, Tsetsos et al. 2012b). A primacy bias whereby early samples carry more sway can be explained if participants believe that samples are drawn in a way that is not conditionally independent (i.e., there is nonstationarity in the generative process even within a trial). For instance, if participants believe that the world changes slowly, then a belief formed early in a trial will be considered a relevant context for evaluating later information, so that early signals will effectively be double-counted and weighted more heavily in the choice. This effect is more prominent when sensory signals are weak but the category of each sample is easy to discern (Lange et al. 2019), with a recency bias potentially prevailing instead in the converse case (Cheadle et al. 2014).

Attractive biases occur in cognitive tasks, where they are often called anchoring effects. Akin to sequential effects, when gambles or risky actions are repeated, participants tend to harbor a superstition that luck comes in streaks, known as the hot hand fallacy (Tversky & Kahneman 1971). Another attractive bias occurs during estimation judgments about facts (e.g., the number of African nations in the United Nations) or prices (of wine or electronics), which were found to be revised upward or downward toward an anchor, an irrelevant number against which the estimate was compared (Arieli et al. 2003, Tversky & Kahneman 1974). In a field study, participants moving to a city from more expensive towns will accept higher rental prices (Simonsohn & Loewenstein 2006). These are analogous to a well-described perceptual effect whereby estimation of a visual feature (such as the motion direction in a field of dots) is attracted in the direction of a previous categorical judgment (Jazayeri & Movshon 2007).

These examples highlight how biases in perception, judgment, and valuation might arise if participants import their beliefs about how the natural world is structured—the relative prevalence of different sensory features, or an assumption that sensory signals are conditioned on the past—into laboratory-based tasks where these assumptions are violated. However, in the context of the model introduced above, these explanations rely on the existence of a memory signal $p(x)$ that arrives independently to bias judgment toward extant beliefs. It is not clear how such a signal might be encoded or retrieved over the very different timescales of expectation for which attractive biases are proposed to operate in the theories described above (over the lifetime, over a context of several minutes or seconds, or over a few hundred milliseconds). Moreover, as outlined below, there is a potentially complementary way that past experience can bias inference, via variation in the encoding function rather than the prior.

4.2. Efficient Coding and Repulsive Choice Biases

Experience with the world shapes the connections among neurons and thus the shape of tuning curves. Consequently, stimulation history can lead to biased perception if it drives asymmetries in the encoding function. The neurons modeled in **Figure 2a** have homogeneous tuning curves; thus, if the prior were flat, then the decoded estimate of *x* would be unbiased. However, experience with natural environments may lead organisms to encode some features or values with higher gain or fidelity, via tuning functions that differ in shape or height (**Figure 2b**).

Distorted encoding functions can lead to biased or suboptimal inference but may be adaptive for agents with finite computational resources. An efficient coding objective can be achieved if neurons are tuned to maximize the mutual information between a stimulus input variable and a neural response variable, minimizing redundancy in neural coding (Attneave 1954, Barlow 1961, Laughlin 1981). With regard to first-order statistics of the environment, this means that expected stimuli are represented with higher gain or fidelity, so that psychometric, neurometric, or econometric encoding functions match the cumulative distributions over expected sensory or value signals. For example, one region of feature space may be encoded by a higher density of neurons, or with neuronal tuning functions that are sharper or that differ in amplitude (Simoncelli 2009). For a model in which each weight could be tuned to satisfy a reconstruction or classification objective (i.e., a neural network), approximations to efficient coding can emerge naturally as a consequence of training (Linsker 1988).

One phenomenon that can be explained in this way is the concavity of utility functions. In the case of magnitude or number, we can explain the form of psychometric or econometric functions under the assumption that encoding is log normal (or otherwise compressive), meaning that the tuning curves for larger quantities are broader, leading to less precise estimates (Woodford 2020). This is reminiscent of the well-known Weber–Fechner law for sensory magnitudes. In fact, refinements of both Weber's theory (Stevens 1957) and Prospect Theory (Kahneman & Tversky 1979) propose a power-law mapping from objective to subjective intensities or values. Decisions about numerical quantities, intervals, or other abstract representations of magnitude seem to display compressive encoding, with higher discriminability for lower values (Dehaene 2003), a phenomenon that can even be observed (in similarity judgments) for symbolic numbers (Shepard et al. 1975). The neural tuning curves proposed by this encoding model (**Figure 2c**) resemble those observed during neural recordings from the dorsal stream cortices of macaque monkeys performing a numerosity judgment task (Dehaene 2003, Nieder & Dehaene 2009).

In combination with Bayesian decoding, a log-normal encoding scheme will ensure that utility functions exhibit the concavity that is characteristic of diminishing marginal utility, and that value estimates are contracted toward the prior mean (Khaw et al. 2018). Of course, the natural statistics of number, value, or quantity may be hard to measure, so we must make some assumptions. For example, it seems plausible that across measurement domains, lower quantities prevail, leading to a lopsided prior—I might purchase a coffee for $2 on my daily commute but consider buying a new laptop for $2,000 only every few years (Woodford 2012). Indeed, the law of anomalous numbers states that first-digit distributions should be logarithmically rather than uniformly distributed, with the number 1 occurring about six times more often than the number 9 (Benford 1938). Another possibility is that the distribution of extant quantities (and thus the prior) peaks at a point midway between "few" and "many," leading small values to be overestimated and large values to be underestimated. Either of these accounts could explain the observed conservative bias in the estimation of quantities or values (Kaufman et al. 1949, Woodford 2020).

It is important to understand that encoding that is distorted under an efficiency principle can paradoxically lead to cases where subjective percepts or values are repulsed away from the

prior, sometimes called anti-Bayesian effects. Where coding density is higher or neurons are more sharply tuned for more prevalent features, a skew emerges in the envelope of the neural population response (or equivalently in the likelihood function) with a long tail away from the peak of the prior (**Figure 2b**). This powerful observation allows us to explain otherwise perplexing data. For example, in orientation perception, the tendency of uncertain tilt judgments toward cardinal or oblique axes varies across studies, with some studies reporting a generalized repulsion of judgment away from cardinal (de Gardelle et al. 2010, Smith 1962, Tomassini et al. 2010). This finding can be explained by a combination of efficient coding and biased decoding, under the assumptions that attraction to the prior (through Bayesian decoding) and repulsion from the prior (through efficient encoding) occur in different measure, depending on the precise form of the noise (see Wei & Stocker 2015 for details). This implementation of efficient coding allows us to derive a general law by which discrimination sensitivity and bias are related across multiple stimulus domains (Wei & Stocker 2017). In the case of tilt, the idea that coding density is heightened for more prevalent features would also predict that perceptual sensitivity should be higher for a cardinal stimulus, which is the canonical finding for orientation discrimination performance (Appelle 1972). It also explains why neurons in macaque V1 preferentially code for cardinal angles (Mansfield 1974), which may be the basis for the higher-amplitude V1 BOLD (blood-oxygen-level-dependent) signals elicited by cardinal versus oblique stimuli (Furmanski & Engel 2000).

When economic stimuli are presented in more tangible form—for example, as pictures of snack food items that a participant might want to eat, rather than numeric values—a related efficiency principle seems to apply, except with a prior that falls in the middle of the value range chosen by the experimenter. Illusions of valuation can be observed whereby item valuation is pushed away from a central prior, by analogy with the oblique effect (Polanía et al. 2019). This repulsion effect leads participants to overestimate that those values are more extreme, such as the price of a snack (or the age of a painting), such that judgments are made more consistently and with greater confidence at extremes of the rating scale than in the middle (Lebreton et al. 2015, Polanía et al. 2019).

5. CONTRAST EFFECTS AND REFERENCE DEPENDENCE

Biases that repel decisions from the local context provided by otherwise irrelevant information occurring proximally in space and time are called contrast effects. The earliest formulations of efficient coding were proposed to explain temporal contrast effects occurring during adaptation to prolonged stimulation. After exposure to visual motion, participants experience an aftereffect whereby a static stimulus appears to move in the opposite direction, known as the waterfall illusion (Barlow & Hill 1963). Similarly, the perceived orientation of a grating is repulsed away from an adapting standard (Clifford et al. 2000). These perceptual biases can be understood with recourse to the underlying neurophysiology. Prolonged or repeated stimulation leads to an attenuation in firing, a phenomenon called adaptation or repetition suppression, occurring via either dampening or sharpening of tuning curves (Kohn 2007). **Figure 3a** shows an example of sharpening (Kohn & Movshon 2004). By attenuating the response to the adaptor in favor of nonadapted features, naïve decoding will lead to a percept that is repulsed away from the adapter. This effect can be observed in the dynamic changes in orientation tuning in V1 observed following adaptation to a tilted grating (Dragoi et al. 2001) (**Figure 3b**). Indeed, neural sensitivities are rescaled at every level of processing—from the sensory periphery to the cortex—to account for the likely range of experienced values (Carandini & Heeger 2012). These dynamics satisfy the principle of efficient coding by sensitizing the system to those changes that inevitably occur after release from adaptation.
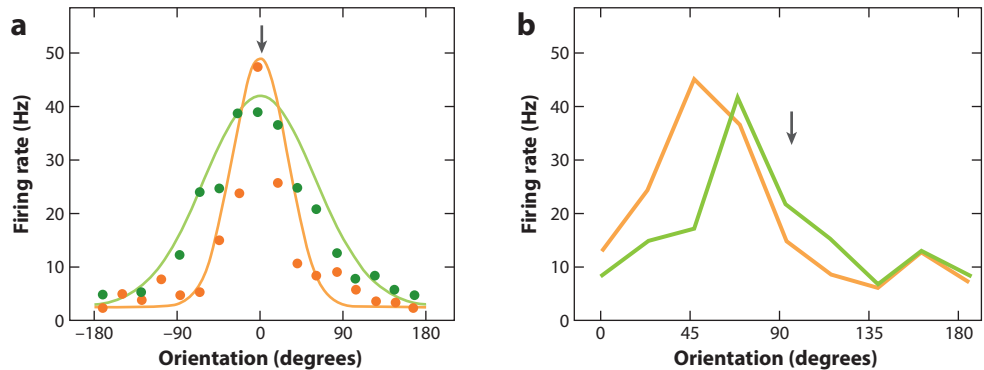
Figure 3

(*a*) Tuning curve for a middle temporal neuron before (*green*) and after (*orange*) exposure to an adapting stimulus at 0° (*arrow*). Dots represent data; lines represent the best-fitting Gaussian tuning curves. (*b*) Tuning curve for a V1 neuron before (*green*) and after (*orange*) adaptation to the stimulus orientation shown by the arrow. Data in panel *a* adapted from Kohn & Movshon (2004). Data in panel *b* adapted from Dragoi et al. (2001).

5.1. Reference Dependence and Normalization

Contrast effects are an example of reference-dependent computation, whereby we evaluate a stimulus as either good or bad relative to the local context. Reference-dependent decisions can even be made about highly abstract values such as moral principles or social prospects. For example, participants deciding appropriate prison sentences for fictitious offences tend to make more lenient judgments in the trial after evaluating a particularly heinous crime (Pepitone & DiNubile 1976), and participants at speed dating events exhibit negative correlation in ratings from one candidate to the next, as if a poor match makes the next seem more attractive (Bhargava & Fisman 2014). However, reference dependence is also a key aspect of Prospect Theory, allowing it to account for framing effects observed in decisions about lotteries under the assumption that providing an initial financial endowment simply shifts the reference, so that overall net gains are in fact perceived to be losses (Tversky & Kahneman 1981). The power of reference-dependent computation is demonstrated by a study that asked participants to pay to avoid electric shocks. The price of pain varied with the overall range of endowments offered in a block, as if participants' valuation were strongly biased by the local context (Vlaev et al. 2009). The relativization of valuation across a temporal context is strikingly apparent in neural recordings from both the orbitofrontal cortex and the parietal cortex, where neuronal signals typically scale with the value of stimuli or actions. In these areas, both firing rates and BOLD signals are rescaled to match the range or average of values that have been recently experienced, consistent with a referent-dependent evaluation (Cox & Kable 2014, Louie et al. 2011, Padoa-Schioppa 2009, Rangel & Clithero 2012).

5.2. Decoy Effects and the Tilt Illusion

Contrast effects can also be driven by spatial context. Consider **Figure 4*a***, in which observers are asked to report the tilt of a grating in the context of flanking stimuli that are irrelevant to the decision. The (horizontal) central grating appears tilted clockwise of horizontal, that is, repulsed away from the context provided by the flanking gratings. This effect, known as the tilt illusion (Blakemore et al. 1970), is related to the illusion shown in **Figure 1*a***. The tilt illusion can be explained via a mechanism related to the repulsive aftereffects described above, except that
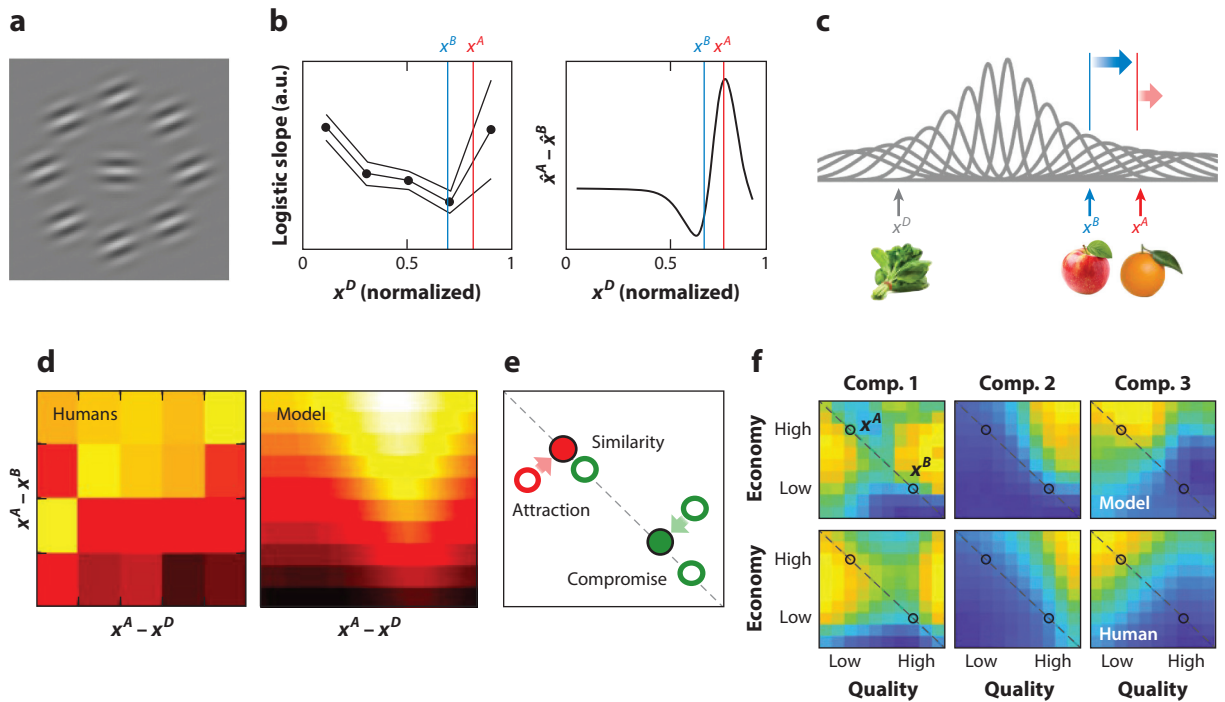
Figure 4

(*a*) The tilt illusion. The central grating appears tilted clockwise relative to its true horizontal orientation, due to repulsion from the flankers. (*b*) The distracter effect. (*Left*) The relative impact of a decoy x^D on propensity to choose the highest-valued option x^A over the next-best item x^B. (*Right*) Model predictions from Li et al. (2018). (*c*) Illustration of how the model produces the decoy effect. As x^D grows in value, value estimates for x^B are repulsed so that they are more similar to x^A, provoking confusion. (*d*) Data from Chau et al. (2014) showing the probability of choosing the x^A as a function of $x^A - x^B$ and $x^A - x^D$, as well as the predictions of the model from Li et al. (2018). (*e*) Illustration of the canonical attraction, compromise, and similarity effects. Open circles represent decoy stimuli; filled circles represent two equally valued targets. (*f*) First three principal components of (*top*) the full map of decoy effects measured empirically in humans, alongside (*bottom*) the predictions of the model. Panel *b* adapted from Louie et al. (2013). Panel *f* adapted from Dumbalska et al. (2020).

attenuated firing for target-tuned neurons arrives via lateral inhibition rather than adaptive mechanisms owing to the biophysics of neural firing (Carandini & Ferster 1997). This inhibitory signal can be thought of as a form of normalization, whereby firing rates evoked by a target stimulus $r(x_i)$ are dampened in proportion to the activity of all neurons $\sum_i r(x)$, including those coding for neighboring features (Carandini & Heeger 2012, Louie & Glimcher 2012). In terms of the model introduced above, spatial contrast effects can occur if neurons sensitive to contextual features become more sharply tuned, for example, due to rapid, first-pass (or gist-based) image processing. The result of this dynamic gain control process is that the target is asymmetrically encoded in the neural population, leading to a repulsive bias in feature estimation away from the context (**Figure 2b**).

In value-guided choices, preferences for a good such as a snack or property rental can depend irrationally on another decoy item that is irrelevant by virtue of being unwanted (lower valued) or unavailable, violating basic principles of decision theory (Luce 1959). For example, in both monkeys and humans, a value-based distracter effect occurs when preference for the higher-valued option *A* over the lower-valued option *B* is first diminished and then accentuated as the lowest-valued decoy *D* grows in value (**Figure 4b,c**). One theory holds that decoy effects for value share

the same computational principles as repulsive biases in perception such as the tilt illusion (Louie & Glimcher 2012). For example, the distracter effect might occur because as D increases in value, it first repulses subjective value estimates of B, making them more similar to A, but subsequently repulses A from B, promoting more choices for the higher-valued option (Li et al. 2018).

Decoy effects have been most intensively studied for choices among multiattribute stimuli, such as consumer goods defined by inverse price (economy) and quality (Busemeyer et al. 2019). Three long-standing decision biases, known as the attraction, similarity, and compromise effects (**Figure 4e**), have been defined in this setting, whereby choices between two isopreferred items (e.g., a more expensive but better-quality product and a cheaper but lower-quality alternative) are biased in the presence of a decoy item. A great deal of research has been directed toward searching for a single, parsimonious explanation for all three biases (Roe et al. 2001). Interestingly, the gain control–based model described above (when elaborated to handle decisions about multidimensional stimuli) offers one such explanation. Repulsion from the decoy context (mediated by sharper tuning for neurons sensitive to context features) and a preference asymmetry for high- versus low-price/quality options are in theory sufficient to account for all three decoy effects (Li et al. 2018). In a subsequent empirical study, participants judged real estate for price and quality while the full-space decoy influence was exhaustively mapped. The results confirmed the intuition that decoy effects have a single, common origin in repulsive contextual biases (Dumbalska et al. 2020). The explanation provided is related to accounts based on normalization, but with an additional nonlinear transformation that captures the way the repulsive bias falls off with distance from the context (Landry & Webb 2021).

5.3. Congruency Effects

How decisions are biased by distracting information has also been studied under the rubric of congruency effects. Where an imperative stimulus is paired with a distracter that is perceptually dissimilar (stimulus conflict) or associated with a competing action (response conflict), decisions are slower and more error prone. Classic examples of this phenomenon occur when participants are asked to name the ink color of color words (Stroop 1935) or discriminate the pointing direction of a central arrow that is flanked by opposing arrows (Eriksen & Eriksen 1974). It is usually assumed that congruency effects occur because of competition among neurons coding for competing actions, which can be mitigated by the voluntary deployment of control processes (Botvinick et al. 2001). However, another perspective is that reaction times differ between congruent and incongruent trials because of contextual facilitation (Polat & Sagi 1994). In other words, flanker effects are related to the tilt illusion, in that processing of a congruent target is facilitated by context-driven gain control process. This permits two interesting predictions. First, varying the noise in the flanking stimuli should affect congruent trials more than incongruent trials, and second, there should exist negative flanker effects, in cases where the target is repulsed onto the discrimination boundary, making decisions more difficult. Evidence for both phenomena has been reported (Li et al. 2018, Rouder & King 2003).

6. CATEGORIZATION, HEURISTICS, AND ROBUST DECISIONS

In this final section, we connect our arguments to an important literature that has proposed heuristics as both descriptive and normative accounts of human behavior. Heuristics are decision policies that deliberately discard information, overtly underfitting the extant data (Gigerenzer & Brighton 2009). We can consider heuristics as solutions to a regression problem in which a dependent measure y depends on multiple predictor variables x with potentially different weight or bearing on

the decision. The decision model discussed above is a special case of this setup in which there is either a single predictor variable or two such variables with equal scaling in the case of the decoy effect (cf. **Figure 4d**).

6.1. Heuristics and Robust Averaging

Heuristic models assume that decision-relevant information is discarded, for example, by falsely assigning all predictors equal weight or by using a single predictor x_1 selected to be the best according to its predictive value, when in fact y depends on multiple attributes. For example, when choosing which apartment to rent, I might use a one-reason heuristic by simply deciding whichever of the candidates is most spacious, or deploy an equal weighting heuristic by binarizing the features (spacious/small, central/suburban) and weighing them all equally in making my choice, irrespective of their perceived importance. The second claim is particularly interesting: namely that, by using a decision policy that actively discards information, I might make my choices more robust to noise or uncertainty, such that my policy performs better for making future predictions (a less-is-more effect; Gigerenzer & Brighton 2009). Although a model that uses all the available information will always make the most accurate predictions in a noiseless world, in the real world inputs are unique: No two sensory experiences are ever exactly alike, so the training conditions (under which model parameters are estimated) and the test conditions (where performance is evaluated) are corrupted by independent sources of variability. In machine learning, where large models are deployed and overfitting to the training data is a major concern, robustness is enhanced by regularization. For example, a regularization method known as dropout allows randomly selected groups of units to participate in learning on each training cycle (Srivastava et al. 2014). In fact, neural networks that are forced to adopt a sort of tallying heuristic by setting all weights to (positive or negative) unity do remarkably well even on complex classification problems (Frankle & Carbin 2019).

Heuristics have been studied largely in the context of predictions about high-level variables, such as salaries, car accidents, city sizes, or outcomes of sports events (Gigerenzer & Gaissmaier 2011), but the principles they invoke are general. For example, under the tallying heuristic only the sign of the predictor is used for decision-making, so that inputs are effectively converted from continuous to discrete signals. This implies an encoding function that dramatically repulses subjective estimates away from a criterion, so that $r(x)$ takes one value for stimuli falling below the threshold and another value above. Interestingly, when observers average the color, shape, or tilt of an array of visual items, they employ a soft version of this discretization principle, behaving as if they are evaluating each item as repulsed from the category boundary. This robust averaging can be observed in the decision weights for each item, which are higher for those that are inlying or close to the bound (de Gardelle & Summerfield 2011, Haberman & Whitney 2010, Li et al. 2017).

Like other policies that rely on distorted encoding or decoding, this (soft) tallying behavior incurs a cost to decision accuracy. However, discretization can confer a robustness on choices, protecting them from uncertainty that might arise downstream—due to noise in memory or integration processes or action selection. Consider the case where the decoder integrates decision information from multiple samples. Discretization will minimize cases where the sum falls close to the bound, and is flipped in sign by the application of (Gaussian) noise (Li et al. 2017). In other words, in the face of (late) noise in the inference process itself—which may prevail for many decisions (Drugowitsch et al. 2016)—it may be better to make decisions about discrete rather than continuous quantities. Related evidence suggests that when participants are asked to make interim decisions during evidence integration, the accumulated information is discretized, presumably for similar reasons (Talluri et al. 2021). This principle can be even be applied to comparisons of lotteries defined by a probability and magnitude of outcomes. Power-law distortions of value, such

as those proposed by Prospect Theory, have the effect of repulsing the difference in utility of the two gambles away from indifference in a precisely analogous manner, and thus are reward maximizing in the face of late decision noise (Juechems et al. 2021). Therefore, this finding provides an explanation for concave utility functions that is potentially complementary to those that appeal to the distribution of values in the natural world, and more research is needed to disambiguate these views (Woodford 2020).

6.2. Selective Integration

A related discretization principle occurs when participants average multiple (uniformly sampled) numbers or features occurring successively in concomitant streams, and make a comparative judgment based on their average. For example, participants might view two parallel streams of numbers and judge which has the higher (or lower) sum (Tsetsos et al. 2012a). This experimental setting prompts participants to display violations of rationality including decoy effects, framing effects, and choice intransitivity (Usher et al. 2019). Critically, this ensemble of effects can be explained by a simple principle, known as selective integration. Selective integration proposes that attention is drawn to the more frame-consistent of the two competing samples (i.e., the larger number if judging which number stream is higher), ensuring that it is updated with higher gain during the integration process. Thus, when integrating a 3 (in one stream) and a 7 (in the other), I will update my decision variable by 7 in the latter case and by $3w$ in the former, where $0 > w > 1$. Doing so has the effect of rendering decisions partly sensitive to the tally of sample ranks, allowing for systematically intransitive choices where aggregate continuous and rank-based (discrete) tallies are placed in opposition (Tsetsos et al. 2016). Critically, however, like robust averaging, selective integration is accuracy maximizing in the face of late noise, because the discretization confers a robustness on decision signals. A similar effect was observed in a related study in which numbers from two categories (e.g., red and green) were averaged in a single, continuous stream, and the level of attentional bias was found to vary with manipulations of the difficulty of integration in exactly the way that maximized reward (Spitzer et al. 2017).

7. DISCUSSION AND OPEN QUESTIONS

Our goal in this article is to understand a broad range of behavioral phenomena through the lens of a small set of normative principles. We focus on two seemingly contradictory effects of decision context, in which percepts or values are distorted by being attracted to or repulsed from a reference. Attraction arises from the imperative to use all the relevant information for a choice, including that from the past, when making decisions (Bayesian inference), and repulsion from a need to maximize the (limited) coding capacity of neural system, by adapting tuning curves to minimize redundancy in information processing (efficient coding). As described above, the joint application of these two principles can have counterintuitive effects on perception and valuation, simultaneously driving inferences toward and away from our expectation.

7.1. Timescales of Expectation

The combination of Bayesian inference and efficient coding allows for powerful, expressive models of choice biases under established normative principles (Polanía et al. 2019, Wei & Stocker 2015). However, it remains unclear what determines the balance of attraction and repulsion in context-based decision-making. Why do we sometimes believe in hot hands but sometimes in a gambler's fallacy—the idea that good luck follows bad—even though they make diametrically opposing predictions about the coming outcome?

One possibility is that timescale matters. For example, repulsive effects (such as aftereffects) may be local, low level, and short lived, whereas the longer-lasting effect of conditional expectation of a stimulus during statistical learning may involve mainly attraction toward the prior. For example, short-term repulsion and long-term attraction may be hallmarks of adaptive effects in binocular rivalry (Chopin & Mamassian 2012). Confusingly, however, there is also evidence for the exact opposite: Sequential biases in orientation discrimination are attractive over short windows, but repulsive over the long term (Fritsche et al. 2020). Another possibility is that the balance of attraction to repulsion is determined by the principal locus of noise. For example, one study of the tilt illusion found that the attractive bias increases as uncertainty was increased by broadening the distribution of tilts in an array being averaged (Tomassini et al. 2010), but that the repulsive bias increases when stimuli were rendered less visible through masking and short presentation times (de Gardelle et al. 2010). Wei & Stocker (2015) argue that their model explains this inconsistency if the latter noise is late and the former early, but it is not clear that this is the case—indeed, varying onset asynchronies of masked stimuli have an impact as early as V1 (Macknik & Livingstone 1998), whereas manipulating the variance of stimulus features affects higher dorsal stream structures (Michael et al. 2015).

More generally, the question of whether decision biases are best thought of as due to distortions in encoding or decoding remains unresolved. Do we perceive the world as it is, but make decisions tactically, in ways that account for our own potential failures of memory and inference? Or are our decision biases driven largely by efficient resource allocation, so that dynamically varying sensitivity is wired into encoding functions for perception and valuation? In a way, these questions reframe long-standing debates in the cognitive sciences that relate to the processing stage at which information is filtered (early versus late) and the degree to which decisions are embodied, that is, primarily in the frame of reference of action versus sensation.

7.2. Neurocognitive Mechanisms

To illustrate the computational basis for the irrational phenomena discussed here, we have focused on a model that makes some gross simplifying assumptions and sidesteps some key questions about both algorithm and implementation. Biased decisions are often proposed to arise from a generic prior distribution, but a clear picture of how this prior is acquired, stored, or retrieved is lacking. In general terms, knowledge is either stored in stable changes in synaptic efficacy between neurons (long-term memory) or maintained in persistent or recurrent activity patterns when neurons exchange information across time periods (short-term memory). Some phenomena, such as attractive choice history biases, may occur because the context of past trials functions like an associative or predictive cue, which is known to drive anticipatory firing in cortical neurons sensitive to shapes (Miyashita 1988) or motion directions (Albright 2012). However, the top-down origins of these anticipatory biases remain murky, despite an assumption that long-term memory circuits may be involved (Kok & Turk-Browne 2018). One promising theme seems to be that cognition is shaped by a stream of samples from long-term memory—including, for example, our memories of the value of different experiences (Shadlen & Shohamy 2016), which in turn has been touted as one mechanism driving the shape of utility functions (Bhui & Gershman 2018). Indeed, when sensory experience is tightly controlled from birth in experimental animals, the stochastic neural patterns observed at rest—which collectively reveal the mean input strength to each neuron—recapitulate the average sensory stimulation history (Berkes et al. 2011). This finding suggests an intimate relationship between memory and prediction, as if our memories are our long-term expectations about the world, but leaves us with only vague hints as to the neural mechanisms that make this possible.

7.3. Conclusions

Apparent deviations from rationality in behavior can often be understood in terms of the history of stimulation that agents have received in the natural world. We have discussed how a surprisingly broad range of psychological phenomena can arise from two relatively simple objectives: to use available information to make good decisions and to maximize the efficiency of neural coding. However, a substantial open challenge remains in understanding these principles at the neural level.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Abrahamyan A, Silva LL, Dakin SC, Carandini M, Gardner JL. 2016. Adaptable history biases in human perceptual decisions. *PNAS* 113(25):E3548–57

Akaishi R, Umeda K, Nagase A, Sakai K. 2014. Autonomous mechanism of internal choice estimate underlies decision inertia. *Neuron* 81(1):195–206

Akrami A, Kopec CD, Diamond ME, Brody CD. 2018. Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature* 554(7692):368–72

Albright TD. 2012. On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron* 74(2):227–45

Appelle S. 1972. Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. *Psychol. Bull.* 78(4):266–78

Arieli D, Loewenstein D, Prelec D. 2003. Coherent arbitrariness: stable demand curves without stable preferences. *Q. J. Econ.* 118:73–105

Attneave F. 1954. Some informational aspects of visual perception. *Psychol. Rev.* 61(3):183–93

Barlow HB. 1961. Possible principles underlying the transformation of sensory messages. In *Sensory Communication*, ed. WA Rosenblith, pp. 217–34. Cambridge, MA: MIT Press

Barlow HB, Hill RM. 1963. Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature* 200:1345–47

Basten U, Biele G, Heekeren HR, Fiebach CJ. 2010. How the brain integrates costs and benefits during decision making. *PNAS* 107(50):21767–72

Benford F. 1938. The law of anomalous numbers. *Proc. Am. Philos. Soc.* 78(4):551–72

Berkes P, Orban G, Lengyel M, Fiser J. 2011. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331(6013):83–87

Bhargava S, Fisman R. 2014. Contrast effects in sequential decisions: evidence from speed dating. *Rev. Econ. Stat.* 96(3):444–57

Bhui R, Gershman SJ. 2018. Decision by sampling implements efficient coding of psychoeconomic functions. *Psychol. Rev.* 125(6):985–1001

Blakemore C, Carpenter RHS, Georgeson MA. 1970. Lateral inhibition between orientation detectors in the human visual system. *Nature* 228(5266):37–39

Bogacz R. 2007. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cogn. Sci.* 11(3):118–25

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. *Psychol. Rev.* 108(3):624–52

Bowers JS, Davis CJ. 2012. Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138(3):389–414

Britten KH, Shadlen MN, Newsome WT, Movshon JA. 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12(12):4745–65

Busemeyer JR, Gluth S, Rieskamp J, Turner BM. 2019. Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends Cogn. Sci.* 23(3):251–63

Caraco T. 1981. Energy budgets, risk and foraging preferences in dark-eyed juncos (*Junco hyemalis*). *Behav. Ecol. Sociobiol.* 8:213–17

Carandini M, Ferster D. 1997. A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science* 276(5314):949–52

Carandini M, Heeger DJ. 2012. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13(1):51–62

Chau BKH, Kolling N, Hunt LT, Walton ME, Rushworth MFS. 2014. A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nat. Neurosci.* 17(3):463–70

Cheadle S, Wyart V, Tsetsos K, Myers N, de Gardelle V, et al. 2014. Adaptive gain control during human perceptual choice. *Neuron* 81(6):1429–41

Chopin A, Mamassian P. 2012. Predictive properties of visual adaptation. *Curr. Biol.* 22(7):622–26

Christian B. 2020. *The Alignment Problem: Machine Learning and Human Values*. New York: Norton

Clifford CWG, Wenderoth P, Spehar B. 2000. A functional angle on some after-effects in cortical vision. *Proc. R. Soc. B* 267(1454):1705–10

Cox KM, Kable JW. 2014. BOLD subjective value signals exhibit robust range adaptation. *J. Neurosci.* 34(49):16533–43

Dasgupta I, Gershman SJ. 2021. Memory as a computational resource. *Trends Cogn. Sci.* 25(3):240–51

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8(12):1704–11

de Gardelle V, Kouider S, Sackur J. 2010. An oblique illusion modulated by visibility: non-monotonic sensory integration in orientation processing. *J. Vis.* 10(10):6

de Gardelle V, Summerfield C. 2011. Robust averaging during perceptual judgment. *PNAS* 108(32):13341–46

Dehaene S. 2003. The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* 7(4):145–47

Dragoi V, Rivadulla C, Sur M. 2001. Foci of orientation plasticity in visual cortex. *Nature* 411(6833):80–86

Drugowitsch J, Wyart V, Devauchelle AD, Koechlin E. 2016. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron* 92(6):1398–411

Dumbalska T, Li V, Tsetsos K, Summerfield C. 2020. A map of decoy influence in human multialternative choice. *PNAS* 117(40):25169–78

Eriksen BA, Eriksen CW. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* 16(1):143–49

Feldman NH, Griffiths TL, Morgan JL. 2009. The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116(4):752–82

Fischer J, Whitney D. 2014. Serial dependence in visual perception. *Nat. Neurosci.* 17(5):738–43

Fleming SM, Thomas CL, Dolan RJ. 2010. Overcoming status quo bias in the human brain. *PNAS* 107(13):6005–9

Franconeri SL, Alvarez GA, Cavanagh P. 2013. Flexible cognitive resources: competitive content maps for attention and memory. *Trends Cogn. Sci.* 17(3):134–41

Frankle J, Carbin M. 2019. The lottery ticket hypothesis: finding sparse, trainable neural networks. arXiv:1803.03635 [cs]

Fritsche M, Spaak E, de Lange FP. 2020. A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife* 9:e55389

Furmanski CS, Engel SA. 2000. An oblique effect in human primary visual cortex. *Nat. Neurosci.* 3(6):535–36

Geisler WS. 2008. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59:167–92

Geisler WS, Ringach D. 2009. Natural systems analysis. *Vis. Neurosci.* 26(1):1–3

Gershman SJ, Horvitz EJ, Tenenbaum JB. 2015. Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349(6245):273–78

Gigerenzer G, Brighton H. 2009. *Homo heuristicus*: why biased minds make better inferences. *Top. Cogn. Sci.* 1(1):107–43

Gigerenzer G, Gaissmaier W. 2011. Heuristic decision making. *Annu. Rev. Psychol.* 62:451–82

Girshick AR, Landy MS, Simoncelli EP. 2011. Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14(7):926–32

Haberman J, Whitney D. 2010. The visual system discounts emotional deviants when extracting average expression. *Atten. Percept. Psychophys.* 72(7):1825–38

Hanks TD, Summerfield C. 2017. Perceptual decision making in rodents, monkeys, and humans. *Neuron* 93(1):15–31

Heng JA, Woodford M, Polania R. 2020. Efficient sampling and noisy decisions. *eLife* 9:e54962

Hollingworth HL. 1910. The central tendency of judgement. *J. Philos. Psychol. Sci. Methods* 7:461–69

Jazayeri M, Movshon JA. 2007. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature* 446(7138):912–15

Jazayeri M, Shadlen MN. 2010. Temporal context calibrates interval timing. *Nat. Neurosci.* 13(8):1020–26

Jones M, Love BC. 2011. Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav. Brain Sci.* 34(4):169–88; discussion 188–231

Juechems K, Balaguer J, Spitzer B, Summerfield C. 2021. Optimal utility and probability functions for agents with finite computational precision. *PNAS* 118(2):e2002232118

Juechems K, Summerfield C. 2019. Where does value come from? *Trends Cogn. Sci.* 23(10):836–50

Juslin P, Olsson H. 1997. Thurstonian and Brunswikian origins of uncertainty in judgment: a sampling model of confidence in sensory discrimination. *Psychol. Rev.* 104(2):344–66

Kahneman D. 2012. *Thinking, Fast and Slow*. London: Penguin

Kahneman D, Slovic P, Tversky A. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge Univ. Press

Kahneman D, Tversky A. 1979. Prospect Theory: an analysis of decision under risk. *Econometrica* 47(2):263–92

Kaufman EL, Lord MW, Reese TW, Volkmann J. 1949. The discrimination of visual number. *Am. J. Psychol.* 62(4):498–525

Kersten D, Mamassian P, Yuille A. 2004. Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55:271–304

Khaw MW, Li Z, Woodford M. 2018. *Cognitive imprecision and small-stakes risk aversion*. NBER Work. Pap. 24978

Kiani R, Hanks TD, Shadlen MN. 2008. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* 28(12):3017–29

Kingdom FAA, Yoonessi A, Gheorghiu E. 2007. The leaning tower illusion: a new illusion of perspective. *Perception* 36(3):475–77

Knudsen E, Wallis J. 2021. Hippocampal neurons construct a map of an abstract value space. *Cell* 184(18):4640–50.e10

Kohn A. 2007. Visual adaptation: physiology, mechanisms, and functional benefits. *J. Neurophysiol.* 97(5):3155–64

Kohn A, Movshon JA. 2004. Adaptation changes the direction tuning of macaque MT neurons. *Nat. Neurosci.* 7(7):764–72

Kok P, Turk-Browne NB. 2018. Associative prediction of visual shape in the hippocampus. *J. Neurosci.* 38(31):6888–99

Kolling N, Wittmann M, Rushworth MFS. 2014. Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* 81(5):1190–202

Landry P, Webb R. 2021. Pairwise normalization: a neuroeconomic theory of multi-attribute choice. *J. Econ. Theory* 193:105221

Lange RD, Chattoraj A, Beck JM, Yates JL, Haefner RM. 2019. A confirmation bias in perceptual decision-making due to hierarchical approximate inference. bioRxiv 440321. **https://doi.org/10.1101/440321**

Laughlin S. 1981. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforschung.* 36(9/10):910–12

Lebreton M, Abitbol R, Daunizeau J, Pessiglione M. 2015. Automatic integration of confidence in the brain valuation signal. *Nat. Neurosci.* 18(8):1159–67

Li V, Herce Castanon S, Solomon JA, Vandormael H, Summerfield C. 2017. Robust averaging protects decisions from noise in neural computations. *PLOS Comput. Biol.* 13(8):e1005723

Li V, Michael E, Balaguer J, Herce Castanon S, Summerfield C. 2018. Gain control explains the effect of distraction in human perceptual, cognitive, and economic decision making. *PNAS* 115(38):E8825–34

Lichtenstein S, Slovic P, eds. 2006. *The Construction of Preference*. Cambridge, UK: Cambridge Univ. Press

Linsker R. 1988. Self-organization in a perceptual network. *Computer* 21(3):105–17

Louie K, Glimcher PW. 2012. Efficient coding and the neural representation of value. *Ann. N. Y. Acad. Sci.* 1251:13–32

Louie K, Grattan LE, Glimcher PW. 2011. Reward value-based gain control: divisive normalization in parietal cortex. *J. Neurosci.* 31(29):10627–39

Louie K, Khaw MW, Glimcher PW. 2013. Normalization is a general neural mechanism for context-dependent decision making. *PNAS* 110(15):6139–44

Luce RD. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley

Macknik SL, Livingstone MS. 1998. Neuronal correlates of visibility and invisibility in the primate visual system. *Nat. Neurosci.* 1(2):144–49

Mansfield RJW. 1974. Neural basis of orientation perception in primate vision. *Science* 186(4169):1133–35

McDermott R, Fowler JH, Smirnov O. 2008. On the evolutionary origin of Prospect Theory. *J. Politics* 70:335–50

McGrayne SB. 2011. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven, CT: Yale Univ. Press

Michael E, de Gardelle V, Nevado-Holgado A, Summerfield C. 2015. Unreliable evidence: 2 sources of uncertainty during perceptual choice. *Cereb. Cortex* 25(4):937–47

Miyashita Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335(6193):817–20

Nickerson RS. 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2:175–220

Nieder A, Dehaene S. 2009. Representation of number in the brain. *Annu. Rev. Neurosci.* 32:185–208

Oaksford M, Chater N. 1994. A rational analysis of the selection task as optimal data selection. *Psychol. Rev.* 101:608–31

Oaksford M, Chater N. 2003. Optimal data selection: revision, review, and reevaluation. *Psychon. Bull. Rev.* 10(2):289–318

Oaksford M, Chater N. 2009. Précis of Bayesian rationality: the probabilistic approach to human reasoning. *Behav. Brain Sci.* 32(1):69–84; discussion 85–120

Padoa-Schioppa C. 2009. Range-adapting representation of economic value in the orbitofrontal cortex. *J. Neurosci.* 29(44):14004–14

Parpart P, Jones M, Love BC. 2018. Heuristics as Bayesian inference under extreme priors. *Cogn. Psychol.* 102:127–44

Pepitone A, DiNubile M. 1976. Contrast effects in judgments of crime severity and the punishment of criminal violators. *J. Personal. Soc. Psychol.* 33(4):448–59

Petzschner FH, Glasauer S, Stephan KE. 2015. A Bayesian perspective on magnitude estimation. *Trends Cogn. Sci.* 19(5):285–93

Polanía R, Woodford M, Ruff CC. 2019. Efficient coding of subjective value. *Nat. Neurosci.* 22:134–42

Polat U, Sagi D. 1994. Spatial interactions in human vision: from near to far via experience-dependent cascades of connections. *PNAS* 91(4):1206–9

Prelec D. 1998. The probability weighting function. *Econometrica* 60:497–528

Ramachandran VS. 1988. Perception of shape from shading. *Nature* 331(6152):163–66

Rangel A, Clithero JA. 2012. Value normalization in decision making: theory and evidence. *Curr. Opin. Neurobiol.* 22(6):970–81

Roe RM, Busemeyer JR, Townsend JT. 2001. Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychol. Rev.* 108(2):370–92

Rouder JN, King JW. 2003. Flanker and negative flanker effects in letter identification. *Percept. Psychophys.* 65(2):287–97

Shadlen MN, Shohamy D. 2016. Decision making and sequential sampling from memory. *Neuron* 90(5):927–39

Shafir E. 1993. Choosing versus rejecting: why some options are both better and worse than others. *Mem. Cogn.* 21(4):546–56

Shepard RN, Kilpatric DW, Cunningham JP. 1975. The internal representation of numbers. *Cogn. Psychol.* 7(1):82–138

Simon HA. 1955. A behavioral model of rational choice. *Q. J. Econ.* 69(1):99–118

Simoncelli EP. 2003. Vision and the statistics of the visual environment. *Curr. Opin. Neurobiol.* 13(2):144–49

Simoncelli EP. 2009. Optimal estimation in sensory systems. In *The Cognitive Neurosciences*, ed. M Gazzaniga, pp. 525–35. Cambridge, MA: MIT Press. 4th ed.

Simonsohn U, Loewenstein G. 2006. Mistake 37: the effect of previously encountered prices on current housing demand. *Econ. J.* 116(508):175–99

Sims CA. 2003. Implications of rational inattention. *J. Monet. Econ.* 50:665–90

Smith SL. 1962. Angular estimation. *J. Appl. Psychol.* 46(4):240–46

Spitzer B, Waschke L, Summerfield C. 2017. Selective overweighting of larger magnitudes during noisy numerical comparison. *Nat. Hum. Behav.* 1:0145

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15(56):1929–58

Steiner J, Stewart C. 2016. Perceiving prospects properly. *Am. Econ. Rev.* 106:1601–31

Stephens DW, Krebs JR. 1986. *Foraging Theory*. Princeton, NJ: Princeton Univ. Press

Stevens SS. 1957. On the psychophysical law. *Psychol. Rev.* 64(3):153–81

Stewart N, Chater N, Brown GDA. 2006. Decision by sampling. *Cogn. Psychol.* 53(1):1–26

Stocker AA, Simoncelli EP. 2006. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9(4):578–85

Stone LS, Thompson P. 1992. Human speed perception is contrast dependent. *Vis. Res.* 32(8):1535–49

Stroop JR. 1935. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18(6):643–62

Summerfield C, Tsetsos K. 2012. Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Front. Neurosci.* 6:70

Sun J, Perona P. 1998. Where is the sun? *Nat. Neurosci.* 1(3):183–84

Talluri BC, Urai AE, Bronfman ZZ, Brezis N, Tsetsos K, et al. 2021. Choices change the temporal weighting of decision evidence. *J. Neurophysiol.* 125(4):1468–81

Thaler RH. 2016. *Misbehaving: The Making of Behavioural Economics*. London: Penguin

Todd PM, Gigerenzer G. 2012. *Ecological Rationality: Intelligence in the World*. Oxford, UK: Oxford Univ. Press

Tomassini A, Morgan MJ, Solomon JA. 2010. Orientation uncertainty reduces perceived obliquity. *Vis. Res.* 50(5):541–47

Tsetsos K, Chater N, Usher M. 2012a. Salience driven value integration explains decision biases and preference reversal. *PNAS* 109(24):9659–64

Tsetsos K, Gao J, McClelland JL, Usher M. 2012b. Using time-varying evidence to test models of decision dynamics: bounded diffusion versus the leaky competing accumulator model. *Front. Neurosci.* 6:79

Tsetsos K, Moran R, Moreland J, Chater N, Usher M, Summerfield C. 2016. Economic irrationality is optimal during noisy decision making. *PNAS* 113(11):3102–7

Tversky A, Kahneman D. 1971. Belief in the law of small numbers. *Psychol. Bull.* 76:105–10

Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185(4157):1124–31

Tversky A, Kahneman D. 1981. The framing of decisions and the psychology of choice. *Science* 211(4481):453–58

Urai AE, de Gee JW, Tsetsos K, Donner TH. 2019. Choice history biases subsequent evidence accumulation. *eLife* 8:e46331

Usher M, Tsetsos K, Glickman M, Chater N. 2019. Selective integration: an attentional theory of choice biases and adaptive choice. *Curr. Dir. Psychol. Sci.* 28(6):552–59

Vlaev I, Seymour B, Dolan RJ, Chater N. 2009. The price of pain and the value of suffering. *Psychol. Sci.* 20(3):309–17

Volz KG, Gigerenzer G. 2012. Cognitive processes in decisions under risk are not the same as in decisions under uncertainty. *Front. Neurosci.* 6:105

von Neumann J, Morgenstern O. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press

Wason PC. 1960. On the failure to eliminate hypotheses in a conceptual task. *Q. J. Exp. Psychol.* 12:129–40

Weber AI, Krishnamurthy K, Fairhall AL. 2019. Coding principles in adaptation. *Annu. Rev. Vis. Sci.* 5:427–49

Wei X-X, Stocker AA. 2015. A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. *Nat. Neurosci.* 18(10):1509–17

Wei X-X, Stocker AA. 2017. Lawful relation between perceptual bias and discriminability. *PNAS* 114(38):10244–49

Wolpert DH. 1996. The existence of a priori distinctions between learning algorithms. *Neural Comput.* 8(7):1391–420

Woodford M. 2012. Prospect theory as efficient perceptual distortion. *Am. Econ. Rev.* 102(3):41–46

Woodford M. 2020. Modeling imprecision in perception, valuation, and choice. *Annu. Rev. Econ.* 12:579–601

Yang T, Shadlen MN. 2007. Probabilistic reasoning by neurons. *Nature* 447(7148):1075–80

Yu A, Cohen J. 2009. Sequential effects: superstition or rational behavior? *Adv. Neural Inf. Proc. Syst.* 21:1873–80