

Optimal utility and probability functions for agents with finite computational precision

Keno Juechems^{*1,2}, Jan Balaguer^{*1}, Bernhard Spitzer^{*3} and Christopher Summerfield^{*1}

¹ Dept. Experimental Psychology, Radcliffe Observatory, Anna Watts Building, Woodstock Rd,
Oxford OX2 6GG, UK

² St John's College, St Giles', Oxford OX1 3JP, UK

³ Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee
94, 14195 Berlin, Germany

*equal contribution from all authors

Corresponding Author information: keno.juechems@psy.ox.ac.uk, jua@google.com,
spitzer@mpib-berlin.mpg.de, christopher.summerfield@psy.ox.ac.uk

Classification

Biological Sciences/Psychological and Cognitive Sciences, Social Sciences/Psychological and
Cognitive Sciences

Keywords

Prospect Theory; utility; uncertainty; computational precision

Abstract

When making economic choices, such as those between goods or gambles, humans act as if their internal representation of the value and probability of a prospect is distorted away from its true value. These distortions give rise to decisions which apparently fail to maximise reward, and preferences that reverse without reason. Why would humans have evolved to encode value and probability in a distorted fashion, in the face of selective pressure for reward-maximising choices? Here, we show that under the simple assumption that humans make decisions with finite computational precision – in other words, that decisions are irreducibly corrupted by noise – the distortions of value and probability displayed by humans are approximately optimal in that they maximise reward and minimise uncertainty. In two empirical studies, we manipulate factors that change the reward-maximising form of distortion, and find that in each case, humans adapt optimally to the manipulation. This work suggests an answer to the longstanding question of why humans make “irrational” economic choices.

Significance Statement

When making economic decisions, humans can evaluate probabilities and magnitudes of outcomes in an idiosyncratic way that can lead to poor decisions. This suggests that the internal functions that map objective quantities onto subjective utilities are nonlinear. Here, we ask why utility functions take this form, making the assumption that human decisions are intrinsically variable (corrupted by noise). Using simulations, we show that the canonical nonlinear form of these functions maximises reward and minimises uncertainty for a noisy decision agent. We demonstrate in two experiments that humans adapt optimally to manipulations of outcome certainty. Thus, our results suggest that observed subjective functions may represent an optimal adaptation within the constraints imposed by biology.

Introduction

Utility theories describe how economic choices are made under risk (1). The expected utility of a risky prospect is a function of its value x and probability p (2). Utility functions characterise the potentially nonlinear transformations that x and p undergo when humans make economic choices, such as deciding among monetary gambles. For example, the subjective expected utility U_i of gamble i offering monetary amount x_i with probability p_i might be described by the function

$$U_i = v(x_i) \cdot w(p_i) \quad [1]$$

where $v(x)$ and $w(p)$ are psychometric transduction functions. Faced with uncertain prospects, human participants will often make choices that fail to maximise expected value (2, 3). For example, many people will prefer \$3000 with certainty over an 80% chance of winning \$4000, even though the uncertain sum has higher expected value. Human preferences also tend to reverse irrationally with irrelevant factors. For example, an agent who prefers the certain sum above might well opt for a 0.2 chance of \$4000 over a 0.25 chance of \$3000 – even though this gamble is identical except for a rescaling of the probabilities by a factor of 1/4. Utility theories typically propose variants of $v(x)$ and $w(p)$ that can capture violations of rationality such as this, allowing researchers to build predictive models of human economic choice (4). However, the resulting models *describe* rather than *explain* the policies that humans adopt when making risky decisions. Here, instead, we seek a normative account of the idiosyncratic forms of the empirically measured functions $v(x)$ and $w(p)$, under the assumption that human decisions are corrupted by irreducible noise in neural computation (5–9).

While the precise form of $v(x)$ and $w(p)$ that best capture human choices remains controversial, there is consensus over several points. Firstly, $v(x)$ is a compressive nonlinearity, such as a (sign-conserving) power law function, that inflects around zero, or the status quo wealth (e.g. eq. 6, Methods). It is often further assumed that $v(x)$ is bounded or rescaled to reflect finite neuronal firing rates. For simplicity, we follow this rationale throughout the main text. Secondly, a number of different forms of $w(p)$ have been proposed, which mostly assert that the probability function approximates an inverse s-shape, i.e. is largely convex but with an initial concavity, often up to a fixed point around $p \sim 0.3$. For example, in one popular model (eq. 7, Methods), the shape parameter γ is empirically found to be on the order of 0.6–0.8, yielding an inverse s-shape (10–12).

Why might the brain employ seemingly sub-optimal distortions, especially when evolution presumably exerted selective pressure for reward-maximising choices? One class of explanation has focussed on $v(x)$, arguing that it evolved to maximise neural coding efficiency in a world where prospects valued at \$1 and \$2 are encountered more frequently than those at \$99 and \$100. A compressive function accentuates differences among the former at the expense of the latter, thus allocating processing resources where they are most likely to be needed, such that the shape of $v(x)$ follows the cumulative distribution over real-world prospects (13). For example, such a nonlinear curve would arise if inferences are made by assessing the value of sampled past experiences from memory, where the non-linear curve would then approximate the cumulative distribution function of these past experiences (14,

15). A different account has focussed on $w(p)$, arguing that its inverted-s shape is an adaptation that overestimates the probability of rare events in order to correct for a bias to overvalue a prospect conditional on its being chosen (9). This latter theory relies on encoding of probability being noisy, but remains silent on the possibility that similar noise corrupts the value function. To date, thus, we currently lack a normative theory that can account for the idiosyncratic form of both $v(x)$ and $w(p)$; in this paper, we seek to offer such a theory.

Our theory rests on two key assumptions. The first is that deliberation is an irreducibly noisy process, and choices are intrinsically variable (16, 17). Secondly, we assume that humans prefer outcomes that are predictable over those that are unpredictable. This assumption is based on a long tradition that emphasises that humans show an aversion to uncertainty (18) and theories that emphasise that entropy reduction is a fundamental goal of animal behaviour (19). We thus assume that the attractiveness of an outcome is computed as follows

$$Y_i = EV_i - \theta \times EV_i \times H_i \quad [2]$$

Where EV_i is the lottery's expected value and H_i is its associated entropy (see eq.5 below). The mixing parameter θ scales the decision-maker's sensitivity to entropy, and where with $\theta = 0$ we simply assume that agents seek to maximise expected value, as assumed by canonical theories.

Our approach assumes that sensory signals (here, symbolic numbers) are encoded with high precision, but that the cognitive process by which probability and value are multiplied and compared is limited by the capacity of memory and attentional processes (we call this limitation "late noise"). Following a long tradition in the cognitive sciences (20), we thus model the probability of choosing gamble 1 over 2 as

$$q(g_1) = \Phi[U_1 - U_2] = \frac{1}{1 + e^{-(U_1 - U_2)/\sigma}} \quad [3]$$

where $\Phi[\cdot]$ is a logistic function with inverse slope σ (although assuming a probit function would yield similar results). The parameter σ scales with the inverse computational precision of human choices: decisions made under higher values of σ are more prone to error. In standard econometric accounts, including those that assume encoding noise, σ is usually treated as a nuisance; it simply allows for random variability in decisions, consistent with the ubiquitously observed sigmoidal form of psychometric functions. We similarly assume here that σ is an inevitable feature of our psychological apparatus, defined as an irreducible "late" noise term or bound on the precision of information processing. We then ask a simple question: given a fixed processing capacity which constrains σ , what is the reward-maximising form of $v(x)$ and $w(p)$? We address this optimisation problem, and in doing so, attempt to derive the minimal assumptions it is necessary to make about the quantity that humans are seeking to optimise during economic choices, in order for the canonically described form(s) of the functions $v(x)$ and $w(p)$ to be optimal, i.e. reward maximising (and/or uncertainty minimising) for the agent.

Results

Consider an agent choosing among a pair of lotteries. Each lottery $[x_i, p_i]$ involves an opportunity to obtain a monetary sum x_i with probability p_i , or otherwise nothing. We denote the binary choice between two lotteries as $[x_1, p_1; x_2, p_2]$. We first consider the case where the mixture parameter θ in eq. 2 is set to zero, so that $Y_i = EV_i = x_i \cdot p_i$ and thus humans should simply be wealth (or reward) maximising, as is commonly assumed in standard models (21). Later, we will show that assuming $\theta > 0$ is necessary to fully capture the canonical value and probability functions. For clarity, we highlight the difference between U and Y : in our formulation U is an internal psychological quantity that is calculated from distorted probability and magnitude of each lottery, whereas Y is defined externally and depends on our assumptions about what is optimal.

The loss function over n choices is then computed as follows:

$$L(U(\mathbf{x}, \mathbf{p}), Y(\mathbf{x}, \mathbf{p})) = -\frac{1}{n} \sum_1^n \Phi[U_1 - U_2] \cdot Y_{1,n} + (1 - \Phi[U_1 - U_2]) \cdot Y_{2,n} \quad [4]$$

Critically, we also assume that agents make decisions with finite precision; in other words, decisions are corrupted by irreducible noise at the level of evaluation or choice, modelled by an inverse slope parameter σ (eq. 3). This conceptualization of noise in the decision-making process has a long tradition (20) and it represents a standard choice across the neurosciences, psychology, and computer science (21, 22).

To begin with, we relax the assumptions that typify classical utility theories and treat this as an unconstrained optimisation problem. We assume a set of choices for which magnitudes x_i (e.g. gamble outcomes in currency units) are drawn uniformly in the range $[0,1]$ and probabilities p_i span the full range $[0,1]$. We define each of j intervals into which x and p may fall, and estimate freely decision coefficients v_j^x and w_j^p for each interval (here, $j = 10$) that minimise the loss term in eq. 4. We solve the optimisation problem separately for different values of $\sigma \in [0:0.02:0.1]$ (see Methods). The resulting coefficients v_j^x (upper panels) and w_j^p (lower panels) for each (fixed) value of late noise σ (columns) are plotted in **Fig. 1a**. We note that for convenience, we restricted the decision coefficients to fall within the range $[0,1]$ to ensure adequate interpretation of the noise term (which scales with the input ranges) and to bring the coefficients in line with the input range (which are scaled to $[0,1]$ throughout the main text, see Methods).

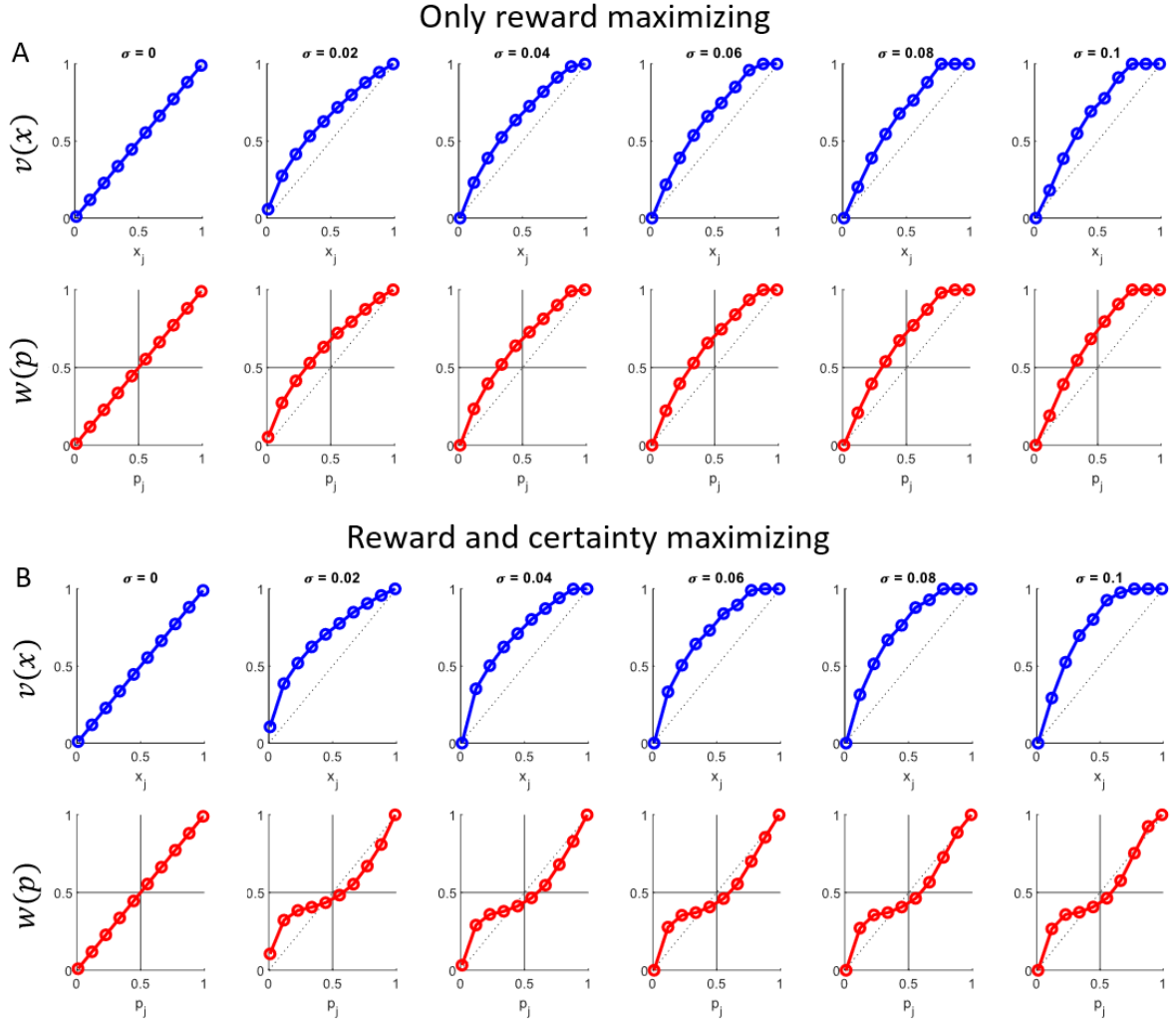


Fig. 1. Optimal value and probability distortions under decision noise. **A.** Optimal (i.e. reward-maximising) values for $v(x)$ and $w(p)$ as derived from eq. 4 under the assumption that $Y_i = x_i \cdot p_i$ (y-axis) plotted against their untransformed counterparts (x-axis) under variable levels of decision noise (columns). Note that for convenience the y-axis was scaled to unity and its values thus do not reflect supra- or sub-linear coding with respect to an ideal observer. **B.** Optimal (i.e. reward and certainty-maximising) values for $v(x)$ and $w(p)$ as derived from eq. 4 under the assumption that $Y_i = x_i \cdot p_i \cdot (1 - H_i)$ (y-axis) plotted against their untransformed counterparts (x-axis) under variable levels of decision noise (columns).

We highlight several features of these simulations. Firstly, in the case where $\sigma = 0$ (leftmost panel of **Fig. 1a**) the reward-maximising policy is to recover parameters v_j^x and w_j^p that are identical to their untransformed counterparts x_j and p_j . This simply means that in the absence of decision noise, expected value can trivially be maximised by multiplying x and p . More interestingly, however, when we assume finite computational precision ($\sigma > 0$) then the policy that maximises expected value is distorted away from the identity line. In fact, both the reward-maximising value function $v(x)$ and weighting function $w(p)$ take the form of a compressive nonlinearity (akin to a power function with exponent $\kappa < 1$), such that an optimal observer will magnify differences between lower magnitudes and between lower probabilities, relative

to their higher counterparts. Thus, a compressive non-linearity of the form presented in eq. 6 provides a compact description of the optimal decision weights. To corroborate this claim, we repeated our simulations under the assumption of power-law transducer functions (eq. 6), again finding a compressive non-linearity that mimics the optimal agent behaviour identified in free-fitting.

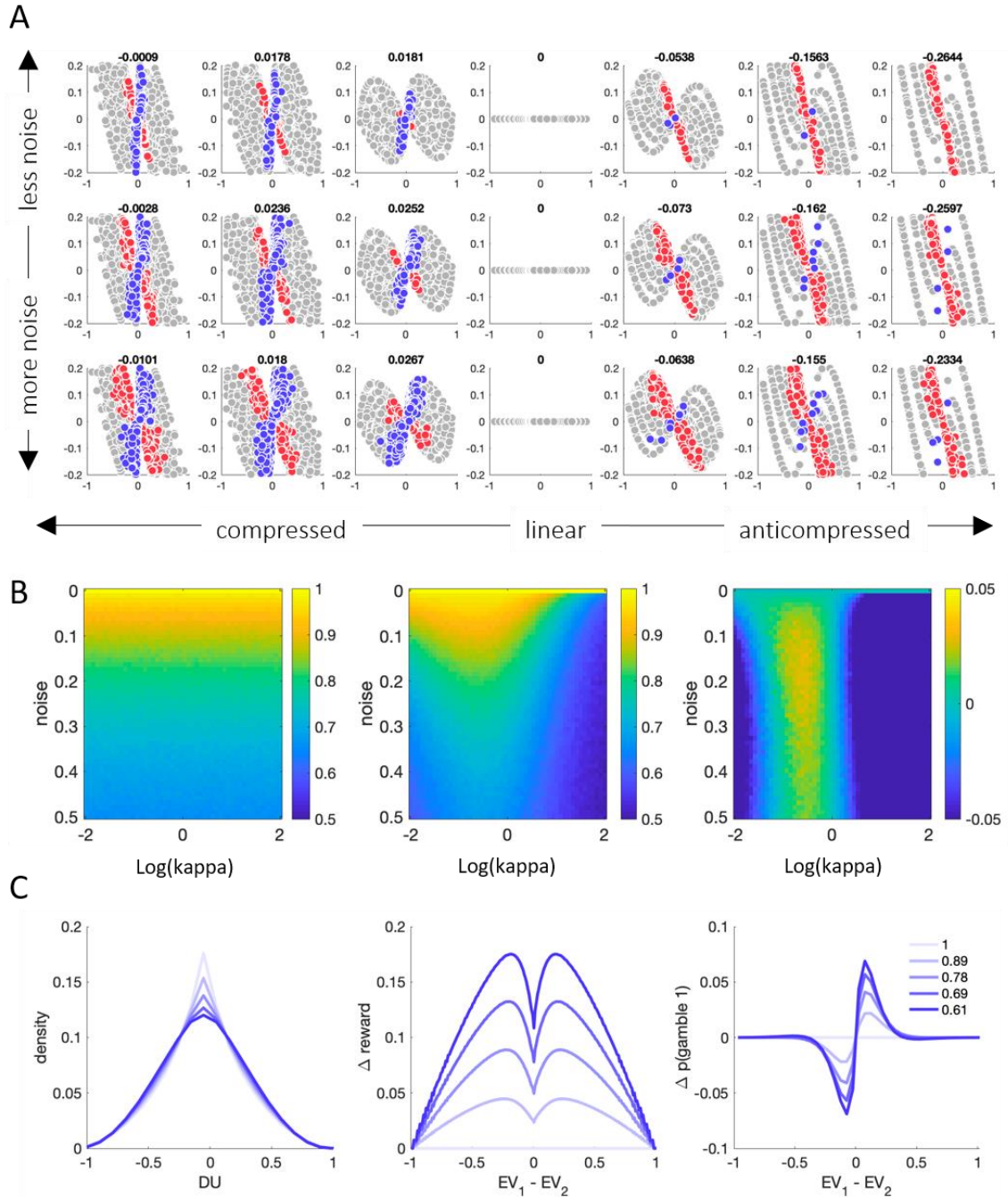


Figure 2. Illustrating optimal value distortion of parametric utility functions. **A.** Each subpanel plots the relative difference in decision utility as a function of the relative expected value of a gamble. Each dot is a unique gamble of the form $[x_1, p_1; x_2, p_2]$. The x-axis denotes $EV_1 - EV_2$ and the y-axis denotes $dv^{lin} - dv^{dis}$ where $dv = U_1 - U_2 + \varepsilon$ under linear (dv^{lin}) and distorted (dv^{dis}) transduction respectively. For $dv^{lin}, \kappa = 1, \gamma = 1$ whereas κ/γ vary from compressive (left columns) to

anticompressive (right columns) under increasing levels of noise (top to bottom; $\varepsilon \in [0.1, 0.3, 0.5]$). Red dots signal those gambles where $\text{sign}(EV_1 - EV_2) = \text{sign}(dv^{lin})$ but $\text{sign}(EV_1 - EV_2) \neq \text{sign}(dv^{dis})$ and blue dots signal the converse, i.e. where $\text{sign}(EV_1 - EV_2) \neq \text{sign}(dv^{lin})$ but $\text{sign}(EV_1 - EV_2) = \text{sign}(dv^{dis})$. The number above each plot indicates the relative fraction of blue dots minus red dots; positive numbers thus indicate that there were more gambles where distortion led to more rewarding choices. **B.** The relative fraction of decision utilities that were of consistent sign with $EV_1 - EV_2$ under linear recoding (left panel), distorted recoding (middle panel), and their difference (right panel), as a function of noise and distortion level (κ). The yellow area shows that compression is reward-maximising where noise is nonzero. **C.** Left: how the relative density of decision utilities changes with different levels of distortion, under various levels of κ . Middle: The expected reward obtained, relative to linear transduction, as a function of $EV_1 - EV_2$ under various levels of κ . Right: The change in probability of choosing gamble 1, relative to linear transduction as a function of $EV_1 - EV_2$, under various levels of κ .

This might seem counterintuitive, but it follows naturally from a consideration of where a limited resource (e.g. selective attention, or neuronal firing rates) can best be allocated to maximise rewards. We illustrate in **Fig. 2** for the case where $Y_i = x_i \cdot p_i$ using the parametric form of the weighting function described in eq. 6. The decision variable on which a choice is based is jointly determined by the difference in the utility of the gambles, i.e. will depend on $+U_1$ and $-U_2$. A compressive value or probability function that distorts U away from Y will increase the probability that the lottery with lower objective expected value will be mistakenly chosen, i.e. distortion increases the number of “sign-flipped” gambles where $p[\text{sign}(U_1 - U_2) \neq \text{sign}(Y_1 - Y_2)]$ (red dots in **Fig. 2a**). Thus, in the absence of decision noise, linear transducers maximise reward. However, an auxiliary effect of the compressive function is to increase the spread of the decision variable $U_1 - U_2$ around zero, including for the majority of cases where the decision sign is not flipped. This ensures that small decision utilities are inflated away from the indifference point, rendering them more robust to decision noise and less likely to result in suboptimal choices (blue dots in **Fig. 2a**). We highlight here that this occurs only where noise in the nervous system is approximately Gaussian or Gumbel in form; if it were (say) uniform, a different pattern would emerge. Due to the repulsion of $U_1 - U_2 + \varepsilon$ away from zero (where ε is a noise term), the decision-maker can increase their expected return by distorting, such that lotteries with relatively small differences incur a relatively higher return under distortion than under a linear (but noisy) encoding (**Fig. 2b-c**). Similar phenomena, including the compressive form of the reward-maximising transducer under “late” decision noise (23), have been reported elsewhere (7, 24).

Whilst the imposition of a compressive nonlinearity has been a hallmark of the utility function $v(x)$ since Bernoulli (25), the compressive form of the optimal probability weighting function shown in **Fig. 1a** (lower panels) is rarely observed (although see the proposal in (26), and the empirical data recorded from rodents in (27) for exceptions). Instead, a more typical form for $w(p)$ is an inverse s-shape with an inflection below the midpoint. However, the curves above were derived under the assumption that humans simply wish to maximise expected value. A long tradition in biology, psychology and machine learning proposes that agents also value information and will pay to reduce their uncertainty about the potential consequences of their actions, for example by choosing predictable over unpredictable outcomes. Presumably, certainty about the future can be beneficial in allowing the animal to forge effective behavioural plans. We build this assumption into our loss term Y , where $\theta > 0$, proposing that

agents will also prefer predictable over unpredictable outcomes (here we refer to the predictability of a gamble as its entropy and not its risk; see discussion below). Specifically, to test whether incorporating uncertainty into the loss function would lead to the standard probability distortion, we set $\theta = 1$ and thus consider a scenario in which the observer evaluates gambles according to a different objective value $Y_i = x_i \cdot p_i \cdot (1 - H_i)$ where the new term H_i is the Shannon entropy of the lottery (with base e):

$$H_i = -p_i \cdot \log(p_i) - (1 - p_i) \cdot \log(1 - p_i) \quad [5]$$

In other words, we assume that human choices seek to maximise reward and minimise outcome uncertainty. Using this new approach, we recomputed the optimal functions in an unconstrained fashion. The results are shown in **Fig.1b**.

As can be seen, under these assumptions we can recover (i) the compressive form of the value function for the domain of gains, (ii) the inverse-s shape of the probability weighting function, and (iii) the inflection around $p \sim 0.3$. These functions clearly resemble the canonical utility functions $v(x)$ and $w(p)$ (10, 28). To verify this contention, we used a model mimicry approach, fitting several classical utility models to these freely derived optimal functions. Under the reward-maximising objective $Y_i = x_i \cdot p_i$ both $w(p)$ and $v(x)$ were well described by a power law function with an exponent of less than one; whereas when risk minimisation was an additional concern, the functions were best described by classical forms of $w(p)$, such as that shown in eq.7 or that proposed by Prelec (29). Together, these simulations thus imply that (i) when decisions are made with finite computational precision (decision noise), distorted utility functions are optimal; and (ii) if we assume that human participants seek to maximise reward and minimise uncertainty, then a distortion of the form predicted by classical utility theory is optimal.

These simulations assume that when humans choose among lotteries, it is deliberation, rather than sensory encoding, that is corrupted by noise. We find it intuitively more plausible that rapidly estimating the product of two numbers should be prone to imprecision than appraising a dollar denoted in Arabic digits. However, previous theories have accounted for the inverted s-shaped form of the probability function under the assumption that encoding, rather than deliberation, is noisy. A natural question, thus, is whether the origin of the noise (encoding vs. deliberation) is a critical factor that determines the form of the reward-maximising value or probability weighting function. In a supplementary section, we show that it does matter; the two models make strikingly different predictions about those parameterisations that maximise reward (**SI Appendix, Fig. S1**). Thus, encoding models, as well as our model (with and without the uncertainty-minimizing assumption, **Fig. 1**) make strikingly different predictions about the optimal distortions.

This observation prompted us to ask whether the distortions in $v(x)$ and $w(p)$ predicted by the model were those displayed by humans under the relevant levels of noise. Specifically, our model allowed us to test a new prediction about human behaviour: near-optimal probability weighting functions should differ according to whether the agent is maximising reward only, or whether she is additionally minimizing uncertainty. Thus, when choices carry no uncertainty, we should expect human behaviour to more closely resemble a model where both $v(x)$ and

$w(p)$ are compressive (noting that in this case, $w(p)$ is not strictly a “probability function”, because x and p are interchangeable). In the following we will term this the “double exponent model”. By contrast, when choices carry uncertainty, we should expect behaviour to resemble that of classical models, where $w(p)$ follows an inverse s-shape – we term this the “Prospect Theory” model. Thus, we should expect the form of the probability weighting function to vary according to whether outcomes are uncertain or not, whereas the value function should not be affected by this manipulation. We note in passing that accounts which suggest that encoding (rather than deliberation) is noisy will not make such a prediction, because sensory stimulation (and thus encoding noise) is held constant between the two conditions, and so this model would predict identical behaviour in the two conditions.

We tested this prediction in two cohorts of human participants ($n = 200$ total) who made incentive-compatible decisions about financial lotteries (in the domain of gains) of the form $[x_1, p_1; x_2, p_2]$. In each of two experiments, one group received the chosen gamble value x_c with probability p_c and zero with probability $1 - p_c$ (uncertain outcome condition, $n = 99$) the other group always received the expected value of the chosen gamble $x_c \cdot p_c$ (certain outcome condition, $n = 101$). In the latter condition, gambles were still described as a magnitude and probability, but of course from the standpoint of a rational agent x and p were strictly interchangeable because the outcome was always the product of these numbers.

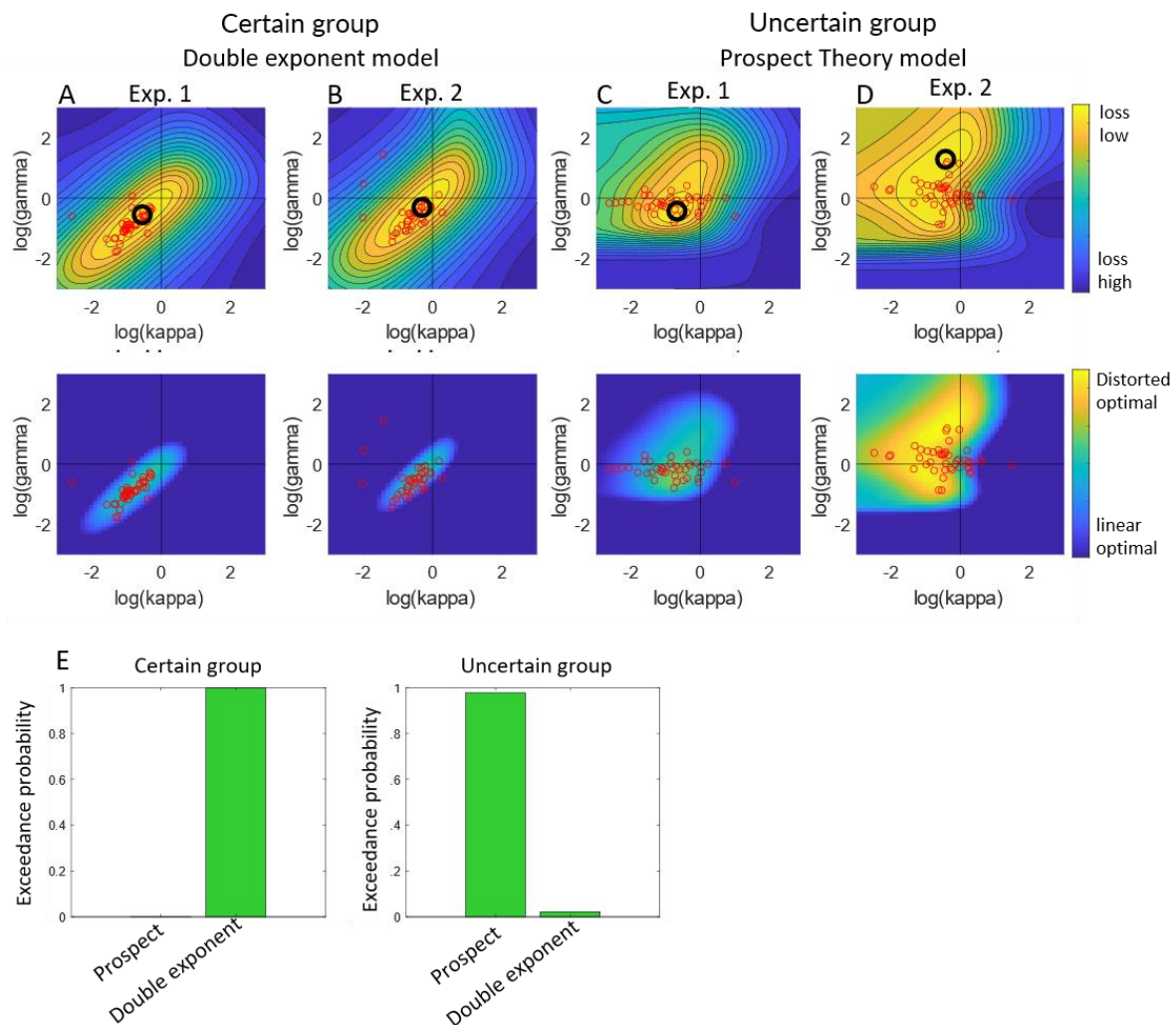


Fig. 3. Upper panels: Loss landscapes showing parameterisations of the winning model in each group that maximise reward (for the certain outcome condition; panels A [Exp.1] and B [Exp.2]) and that maximise reward/minimise risk (for the uncertain outcome condition; panels C [Exp.1] and D [Exp.2]). Warmer colours signal parameterisations that are closer to optimal, and the black circle shows the maximum. Red circles show parameter estimates for individual human participants in each experiment (panel A-B: double exponent model; panel C-D: PT model). Note how humans cluster in the same quadrant as the maximum; see statistics below. Lower panels: the difference in loss between the distorted model and a linear model, i.e. a model with the same estimated noise but with $\log(\kappa) = 0$ and $\log(\gamma) = 0$. Red circles are individual participants. Warmer colours show regions where the distortion increases return over the linear model. **E.** Exceedance probabilities for the Prospect Theory model and double-exponent model in the certain outcome condition (left panel) and uncertain outcome condition (right panel).

In Experiment 1, x and p were drawn randomly for each gamble, such that both lotteries were uncertain. Experiment 2 served as a robustness test in which only one gamble was uncertain, while the other had a probability of one. Sampling was otherwise matched to Exp.1 (see methods). Our model predicts that $w(p)$ and $v(x)$ (under the boundedness assumption) will both resemble power law functions (the “double-exponent model”) in the certain outcome condition, but by classical utility theory in the uncertain outcome condition, even though the expected value distribution is matched between certain and uncertain outcome groups. This is exactly what we found (**Fig. 3e**). Using Bayesian model selection (30) on cross-validated model fits to compare model fits of double-exponent and Prospect Theory (PT) models to human data, we found that the former fit the data best in the certain outcome condition (exceedance probability $[xP] > 0.99$) and the latter fit better in the uncertain outcome condition ($xP > 0.95$).

We conducted our next analysis under the assumption that the double-exponent and Prospect theory (PT) models provided a good reduction of the human policy in the certain outcome condition and uncertain outcome conditions respectively. Both models assume that $v(x) = x^\kappa$ (in the domain of gains explored here) but make different assumptions about $w(p)$: the double exponent model assumes that $w(p) = p^\gamma$ (because x and p are interchangeable) whereas PT assumes the probability weighting function shown in eq.7. This allowed us to plot the (negative) loss landscape $-\mathcal{L}$ under each parameterisation (defined by κ and γ) for a theoretical agent exhibiting the mean level of decision noise estimated across the cohort (**Fig. 3a-d**) and to compare it to the best-fitting parameters for human participants. The surfaces for the certain outcome condition (**Fig. 3a-b**) confirm that the reward-maximising policy for an agent with the same average noise as our participants is to transduce x and p with compressive functions, so that the highest return is obtained for $\log(\kappa) < 0$ and $\log(\gamma) < 0$ (warm colour shading; black circle is the maximum given human levels of noise). For the uncertain outcome conditions (**Fig 3c-d**), the optimal value of $\log(\kappa)$ was always negative but the optimal value of γ varied with experiment: for Exp.1, rewards are maximised with $\log(\gamma) < 0$ but for Exp.2, $\log(\gamma) > 0$ is optimal (note that in the case of the uncertain outcome condition we use the term “optimal” to refer to those solutions that jointly maximise reward and minimise uncertainty, as proposed above). This pattern of predictions closely resembles what was observed from human participants. In each case, the human parameters (red circles in Fig.3) fall in the area that minimises the relevant objective function. Moreover, as can be seen in the

lower panels, the human parameters were those for which the expected return from distorted transducers exceeds the expected return from a linear transducer to the greatest extent.

These intuitions were confirmed by statistical comparison. Parameters $\log(\kappa)$ fell reliably below zero in both conditions of both experiments (Exp 1 Certain Outcome: $t = 12.5$; Exp 2 Certain Outcome: $t = 8.41$; Exp 1 Uncertain Outcome: $t = 7.17$; Exp 2 Uncertain Outcome: $t = 3.87$; all p -values < 0.001). Parameters $\log(\gamma)$ fell reliably below zero in both Certain Outcome conditions (Exp 1: $t = 11.6$; Exp 2: $t = 6.67$; both p -values < 0.001). However, in the Uncertain Outcome condition $\log(\gamma)$ was significantly below zero in Exp 1 ($t = 4.2$, $p < 0.001$) and above zero in Exp2 ($t = 2.46$, $p < 0.02$). These results are consistent with model predictions in each of the cases tested.

A further comparison of the human and reward-maximising value and probability functions is shown in **Fig.4**. Here, we use the unconstrained model to freely estimate the reward-maximising coefficients for each participant individually. The optimal coefficients vary across the cohort as individuals differ in their levels of estimated internal noise, and in the sample of lotteries they viewed. We plot the distribution of coefficients obtained across the cohort for $v(x)$ [blue shading] and $w(p)$ [red shading] and superimpose on top the best-fitting value and probability functions from the double exponent model (certain outcome condition) and Prospect Theory (uncertain outcome condition). As can be seen, human distortions resemble the optimal solution qualitatively in each case tested, although they are slightly weaker than optimal in the uncertain outcome condition.

The thesis advanced here is that distortions in the subjective representation of value and probability are reward-maximising under “late” or decision noise. Our theory thus makes predictions about the relationships among observed parameters themselves, and how they relate to performance. Firstly, because the theory predicts that distortions have a common origin in reward-maximisation (at least in the certain outcome condition), the parameters κ and γ should be correlated: those participants with greater distortion for value should also have greater distortion for probability. Empirically, this is what we found in the certain outcome condition (both r -values > 0.3 , both p -values < 0.001) but not the uncertain outcome condition (both p -values > 0.1). This follows from the assumption that in the uncertain outcome condition κ controls compression of value in the service of reward maximisation, whereas γ largely reflects sensitivity to uncertainty by controlling the inverse s-shape of the probability weighting function in eq. 7.

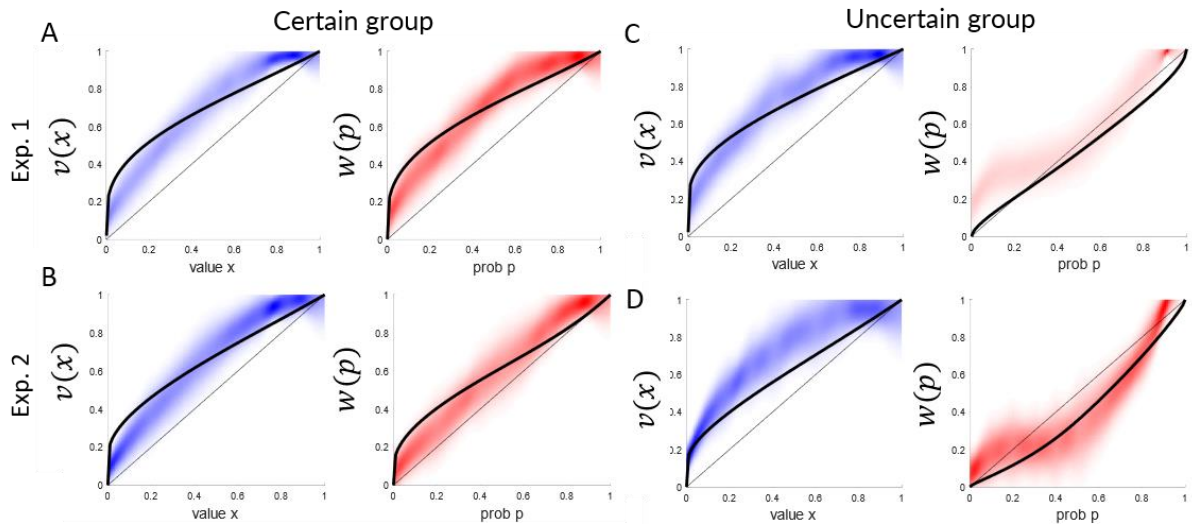


Fig. 4. Each panel shows the reward-maximising form of the functions $v(x)$ (blue shading) and $w(p)$ (red shading). Each function is expressed as a density over optimal estimates derived from each participant in the certain outcome condition (A [Exp.1] and B [Exp.2]) and the uncertain outcome condition (C [Exp.1] and D [Exp.2]). Optimal estimates vary from participant to participant because of distinct noise levels and variation in lottery sampling. Superimposed on each reward-maximising function is the form of the distortion that best fit human choices (estimated from median parameters, shown on each plot). This was estimated from the double-exponent model for the certain outcome condition (A and B) and from Prospect Theory for the uncertain outcome condition (C and D).

Discussion

The nature of the internal representations that guide economic choices has been a question of longstanding interest for psychologists, economists and neuroscientists. Here, we shed light on this question by taking a normative rather than a descriptive perspective. We asked why humans behave as if they distorted estimates of probability and magnitude when making economic decisions. We suggest that they do so because in the presence of decision noise, distorted functions yield higher reward (and lower uncertainty) than undistorted functions. Moreover, we show that the precise form of the distortions commonly observed in economic choice tasks (e.g. where humans make binary choices among lotteries) can be explained under three simple and (we hope) uncontroversial assumptions: (i) human deliberation occurs with finite computational precision; (ii) humans wish to maximise expected value; (iii) humans wish to maximise information gain, i.e. to minimise the uncertainty (or entropy) of a choice. Our empirical work shows that this account explains the effects of switching between otherwise equivalent safe and risky outcomes and can even explain a reversal in the shape of the (probability) weighting function that occurs when one sum is offered with certainty (Exp 2) or not (Exp 1).

We note that our explanations for the form of the utility functions differ sharply from those proposed previously. One common view is that decisions are irrational because observers have a preference for “fast and frugal” computation, i.e. they are willing to sacrifice accuracy for speed when evaluating decision-relevant information (31). This view offers a generic motivation for suboptimal decisions but does not explain the stereotyped form of $v(x)$ and

$w(p)$ observed in empirical studies. With regard to the value function, a compressive nonlinearity has typically been interpreted as implying an aversion to risk, or outcome variance (rather than uncertainty) because it accounts for the oft-observed preference for lottery $[x, 1]$ over $[\frac{x}{k}, k]$ where $k < 1$, i.e. the preference for certainty equivalents that are matched to the expected value of a risky gamble. One could thus assume that the value function has evolved to be compressive because risky choices are generally detrimental in natural environments, where they invoke the possibility of resource depletion and death (32). Critically, however, at other times risky behaviours can be actively beneficial for survival, for example when birds need to harvest critical levels of food in order to survive the night (33, 34). This context-dependence of risk attitudes suggests that they may be secondary to other objectives an organism may harbour, rather than an intrinsic part of its biological objective. Our model does predict aversion to risk, but it does so as a by-product of the need to maximise reward under decision noise with a nonlinear value function. Thus, we caution against interpreting the curvature of the value function and its associated parameter κ as reflecting a parameterized estimate of risk aversion. Indeed, making this assumption in conjunction with a logistic choice function (as in our model) can imply nonmonotonicities in the value function that may be implausible (35). However, it remains an open question whether humans and other animals are additionally averse to risk, over and above the policy implied by the curvature of the value function under late decision noise.

Our model rests on the assumption that noise acts late (i.e. at the decision stage) and can be modelled with a logistic function. Whilst the former assumption is crucial, the latter is not – our results would no doubt generalise to other choice functions of similar form, such as the probit function. However, we also acknowledge that there are cases where noise of the logit/probit form may give rise to incongruous findings (35–37) and note that other choice functions would lead to results that differ from those reported here.

Thus, the explanation we offer also differs from other accounts based on “encoding noise” or efficient coding, although we share their assumption that neural signals are intrinsically imprecise, and that information processing evolved biases which render decisions robust to noise. One class of model argues that humans harbour priors over distributions of likely dollar values (13); because low-valued prospects occur more frequently in natural environments, humans internally model them with increased neural resources and thus higher precision. This can serve to optimize the finite (e.g. binary) coding range available to the neuronal population (38, 39), and to shield the coding scheme from imprecision arising from finite sampling (14, 15). Like risk-based accounts above, this explains the compression to the value function, but is silent about distortions to the probability function, where low-probability events are by definition inversely likely as their high-probability counterparts, making it very hard to formulate an equivalent view that explains nonlinear $w(p)$.

Another prominent model argues that an inverted s-shaped probability distortion would result from noisy probability encoding. In the face of encoding noise, agents will tend to overvalue the option they prefer, a phenomenon often known as the “winner’s curse”; distortions to the probability function that overweight rare events (which are marginally less likely to be valuable) can correct for this bias (9). Our explanation for the distorted probability function differs from that in ref 9. We also note that the argument made in that paper, albeit compelling, would not account for the discrepancy observed here in human behaviour under certain and uncertain

outcomes, because in both cases the probability values will be corrupted by equivalent sensory encoding noise, leading to equivalent results in either condition.

Humans and other animals find states of uncertainty aversive, and seek to avoid them. We explain the distortion to the probability function under the assumption that in addition to their attempt to maximise reward, humans intrinsically seek to minimise the entropy of a decision, that is, the uncertainty about whether it will play out in their favour or not (40, 41). Such entropy is maximised for decisions with fully ambiguous outcomes (e.g. a 50:50 gamble) but in conjunction with the compressive nonlinearity that maximise reward, this yields precisely the inverted-s shaped form to $w(p)$ with an inflection at $p \sim 0.3$. We justify this assumption on two grounds. Firstly, we appeal to a vast literature in the behavioural and cognitive sciences that has emphasised the value of information and the cost of uncertainty (40, 42, 43). When an outcome is known, then plans for the future can be formed; without knowledge of the next outcome, sequential decisions become impossible. In machine learning and neuroscience, unified theories have been built around the assumption that reducing uncertainty is the *raison d'être* of biological organisms (44–48). Secondly, the two empirical studies described here showed that distortions to the probability weighting function critically depend on whether outcomes were certain or uncertain. In other words, when lottery outcomes are equivalently valued but certain, the observed form of $w(p)$ is a compressive nonlinearity, just like $v(x)$ – exactly as predicted by our model. As our formulation of the loss function described in eq. 2 allows further parameterization, our model can also be extended to include uncertainty-seeking behaviour (when setting $\theta < 0$). More generally, θ can be used as a hyper-parameter in future studies to first establish the optimal trade-off between value and uncertainty, and then to use it to generate hypotheses about the resulting shape of the value and probability functions.

Our assumptions about the “canonical” form of the human utility functions draws principally on Prospect Theory. This is not meant to imply a commitment to the specific form of $w(p)$ proposed by eq.7. In fact, very similar results are obtained with the two-parameter forms of the probability weighting function described elsewhere (9). However, we found that the extra flexibility permitted by these functions did not improve the cross-validated model fit to human data, and so we chose to focus on the canonical function proposed by Prospect Theory. We similarly do not seek to endorse Prospect Theory as an ideal observer given the noise and entropy assumption, but simply claim that *if* participants were using Prospect Theory these simple assumptions allow us to predict their functional parameters. Hence, there may be other models which lead to better quantitative fit to our simulated functions. Indeed, the Prospect Theory curves averaged across participants in the uncertain group exhibit some quantitative misfit (Fig. 4b,d).

Moreover, the focus of our paper is on distortions of probability and value in the domain of gains. Whilst an identical argument would explain the mirror-symmetric compressive form of the value function in the domain of losses, the theory proposed here does not explain why losses “loom larger” than gains, having disproportionate influence on choices and further contributing to participants’ failure to maximise expected value (although see recent review articles that question the extent or replicability of loss aversion (49, 50)). Nor does our theory consider the most distinctive contribution of Prospect Theory – the intuition that computation of value is reference-dependent, with all utilities evaluated relative to a status quo given by

the current context. The normative properties of such reference-dependence, for example in the context of efficient coding and efficient computation, have been discussed elsewhere (51).

Author Contributions

BS conceived of the project idea, JB, BS, and CS conceived of the simulation and experiment idea, JB and KJ collected the data, all authors carried out the stimulations, analysed the experimental data, discussed and interpreted the results, and edited the manuscript.

Competing Interests

The authors declare no competing interest.

Acknowledgments

We would like to thank the members of the Summerfield lab, Konstantinos Tsetsos, as well as two anonymous reviewers for their helpful suggestions. The project NEUROABSTRACTION has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 725937).

Methods

Simulations

We obtained the optimal shape of the value and probability weighting functions under decision noise using a model-free approach. We first created all possible combinations of sets of two lotteries of the form $[v, p; 0, 1-p]$ where p and v were drawn from one of $j = 10$ bins of equal size and spacing within the range $[0.01, 0.99]$. These $N = 10,000$ samples were randomly split into training and test sets, after discarding trivial pairs where one lottery had both greater probability and value than its competitor. Next, we asked what decision weight an optimal agent should apply to each bin of value and probability in order to minimize one of two loss functions described below in the training set. Optimization was carried out via gradient descent using Matlab's `fmincon` function, with parameters for each bin initialised to x_j and p_j . We varied the amount of noise in equation (3) by adjusting the parameter σ between 0 and 0.1 in intervals of 0.02, where higher values of σ indicate stronger noise, and repeated the simulation for each level of σ . Having obtained the best fitting estimates we evaluated cross-validated fit on the held-out test set. For convenience and to allow direct comparison, values were scaled to fall within the interval $(0, 1)$. Note, however, that the results of our simulations hold regardless of value scale. Where we used the entropic loss function (eq.2, with $\theta = 1$), the logarithm was base e .

Human Experiments

Ethical approval.

All participants gave informed consent to participate in the study and were free to withdraw at any point. The study was approved by the ethics board of the Medical Sciences Division, University of Oxford (R50750/RE001)

Participants

We recruited $n = 100$ participants online for each of the two experiments on Amazon Mechanical Turk (MTurk). In experiment 1, 59 participants were male and 41 were female. In experiment 2, 62 were male, and 38 were female. Age was only assessed to decade resolution, with most participants falling within the 21 to 30 years range (74% in both experiments). Participants were remunerated for their time with \$5 plus a bonus equivalent to the value of their chosen lottery of one randomly selected trial (scaled to \$0-\$10). We included only those participants whose performance differed significantly ($p < 0.001$) from random as determined by a binomial test ($n = 167$).

Task

Participants performed the task online in their browser and were required to run it in full screen mode for the duration of the experiment. On each of the 250 trials, participants saw two lotteries to the left and right of the centre of the screen and were asked to indicate which of two lotteries they preferred. Participants were only instructed about the visual components of the task and that one lottery would be chosen at random at the end as their bonus, but received no further instructions about how to choose between the lotteries. During the

response window of 20s, both lotteries remained on the screen until participants indicated their response using the left or right arrow key. A dial at the top of the screen indicated the time remaining within the trial. After participants pressed a button, the chosen lottery remained on screen for 1s and was then replaced with the outcome feedback for a further 2s. The next trial began immediately after this. Throughout the experiment, the time elapsed and trials completed were displayed at the top of the screen. Participants took on average 2.38s in experiment 1 and 2.03s in experiment 2 to respond, leading to a total average time for task completion of 23.6 minutes in experiment 1 and 21.5 minutes in experiment 2. There were no breaks during the experiment.

Stimuli

Lotteries consisted of one probability cue expressed as a percentage and one value cue expressed as dollar amount. This indicated that participants had the chance to win \$X with Y% or \$0 otherwise. For convenience, the shared \$0 outcome across lotteries was not displayed on screen. Whether probabilities were displayed above values or vice versa was assigned randomly for each participant. Values and probabilities were sampled from a uniform distribution between \$1-\$99 and 1%-99%, excluding trivial samples where one lottery was better on both value and probability than its competitor (these trivial gambles were also excluded from our simulations). Note, however, that the results obtained from our simulations are not affected by this constraint. Experiment 2 changed this to always include one certain lottery with 100% probability, with the remaining probability and values sampled from the same uniform distribution and subject to the same constraint. In all experiments, participants completed 250 trials and had to indicate their choice by pressing the left or right arrow key.

Feedback

Our main manipulation between groups pertained to how feedback was given to participants. The “certain outcome” condition received as feedback the product of value and probability, in other words the lottery’s expected value, whereas the “uncertain outcome” group received feedback depending on the lottery’s probability: either its dollar value (with probability p) or \$0 (with probability 1-p). Note that the optimal strategy in the absence of noise is always to multiply the value and probability regardless of feedback. Hence, a noiseless ideal observer would not differ in its behaviour between conditions.

Model fitting

Similar to our simulations, we first assessed participants’ weighting functions using a model-free approach. For this purpose, we again binned values and probabilities into $j = 10$ bins and fit one decision weight to each bin using the genetic algorithm. This allowed us to compute the density of decision weight estimates as illustrated in Figure 4, without having to assume a specific parameterized functional form. We assessed how well economic decision-making models accounted for participants’ choices using maximum likelihood fitting. The two models we used were:

First, a standard model from the Prospect Theory (PT) family of models of the following form:

$$v(x) = x^{\kappa}$$

[6]

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \quad [7]$$

Note that we re-scale values and probabilities to the interval [0,1], such that the resulting $v(x)$ and $w(p)$ functions are implicitly bounded within the same interval.

Second, we assumed that participants in the certain outcome group were better fit by a double exponent model as our simulations indicated no difference between the probability and value weighting functions for the return-maximizing agent. Thus, the double exponent model assumed both the $v(x)$ and $w(p)$ weighting to be of the form described in equation (6), albeit with different parameters: κ for $v(x)$ and γ for $w(p)$, employing the same naming convention as in equations (6-7).

Models were fit to the data using a hierarchical model and optimized using the Expectation-Maximization algorithm following software provided in (52). The model first draws a number of parameter samples from a group prior distribution and assesses fit for each sample. The group prior is then updated to better reflect those samples that more accurately predicted behaviour. These two processes proceed iteratively until convergence (i.e. no further improvement to fit is observed by adjusting the group prior or increasing the number of samples). We used the standard normal distribution as group priors. Data were fit separately for each group and experiment.

Model comparison

We compared models using the Variational Bayesian Analysis toolbox (53). We employed a random effect analysis, using the models' log likelihoods to compute the exceedance probability that a given (crossvalidated) model fit participants' data better than all other models. This procedure calculates the likelihood that a given model is more frequently the best model (across participants) compared to all others within the set. This produces a more nuanced model comparison metric than comparisons based on overall, fixed effects model fits (e.g. Bayesian Information Criterion).

Loss landscapes

Having established which of the two models best fit participants' choices in each group and experiment, we asked whether the best fitting parameters fell within the range of optimal parameters for a given model. For this purpose, we plotted the loss landscape of each model over a range of parameter values, evaluating performance on the exact gambles given to participants. We fixed the noise in the simulation to the mean across participants and derived for each parameter pair $[\kappa, \gamma]$ how the model fared relative to a linear agent with parameters $[1, 1]$. As both models used power functions, we explored the parameters in log-space (with base e), where a value of 0 indicates a linear mapping from objective to subjective magnitude. We linearly sampled $n = 50$ points within the range of $[-3, +3]$ in logspace and plotted the value of the loss function (either return-maximizing or entropic) at that point, **Fig. 3**.

Data and code availability

All data, and code to reproduce all figures and to run the simulations are freely available on our lab GitHub repository at this URL: https://github.com/keno-juchems/PT_optimality

705

706

References

1. C. Starmer, Developments in nonexpected-utility theory: The hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* **38**, 104–147 (2000).
2. D. W. Harless, C. F. Camerer, The Predictive Utility of Generalized Expected Utility Theories. *Econometrica* **62**, 1251–1289 (1994).
3. C. F. Camerer, “Prospect Theory in the Wild: Evidence from the Field” in *Choices, Values and Frames*, D. Kahneman, A. Tversky, Eds. (Cambridge University Press, 2000).
4. D. Kahneman, A. Tversky, Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47**, 263–292 (1979).
5. C. Findling, V. Skvortsova, R. Dromnelle, S. Palminteri, V. Wyart, Computational noise in reward-guided learning drives behavioral variability in volatile environments. *Nat. Neurosci.* **22** (2019).
6. J. Drugowitsch, V. Wyart, A. D. Devauchelle, E. Koechlin, Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality. *Neuron* **92**, 1398–1411 (2016).
7. K. Tsetsos, *et al.*, Economic irrationality is optimal during noisy decision making. *Proc. Natl. Acad. Sci.* **113**, 3102–3107 (2016).
8. S. Bhatia, G. Loomes, Noisy preferences in risky choice: A cautionary note. *Psychol. Rev.* **124**, 678–687 (2017).
9. J. Steiner, C. Stewart, Perceiving Prospects Properly. *Am. Econ. Rev.* **106**, 1601–1631 (2016).
10. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).
11. C. F. Camerer, T. H. Ho, Violations of the betweenness axiom and nonlinearity in probability. *J. Risk Uncertain.* **8**, 167–196 (1994).
12. G. Wu, R. Gonzalez, Curvature of the probability weighting function. *Manage. Sci.* **42**, 1676–1690 (1996).
13. M. W. Khaw, Z. Li, M. Woodford, Risk Aversion as a Perceptual Bias. *NBER Work. Pap. Ser.*, 45 (2017).
14. R. Bhui, S. J. Gershman, Decision by sampling implements efficient coding of psychoeconomic functions. *Psychol. Rev.* **125**, 985–1001 (2018).
15. N. Stewart, N. Chater, G. D. A. Brown, Decision by sampling. *Cogn. Psychol.* **53**, 1–26 (2006).
16. L. T. Hunt, What are the neural origins of choice variability? *Trends Cogn. Sci.* **18**, 222–224 (2014).
17. A. Renart, C. K. Machens, Variability in neural activity and behavior. *Curr. Opin. Neurobiol.* **25**, 211–220 (2014).
18. D. Ellsberg, Risk, Ambiguity, and the Savage Axioms. *Q. J. Econ.* **75**, 643–669 (1961).
19. K. Friston, Life as we know it. *J. R. Soc. Interface* **10** (2013).
20. R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis* (Wiley, 1959).
21. R. S. Sutton, A. G. Barto, *Reinforcement Learning*, 2nd Ed. (MIT Press, 2018).
22. R. C. Wilson, A. G. E. Collins, Ten simple rules for the computational modeling of behavioral data. *Elife* **8**, 1–33 (2019).
23. V. Li, S. Herce Castañón, J. A. Solomon, H. Vandormael, C. Summerfield, Robust averaging protects decisions from noise in neural computations. *PLoS Comput. Biol.* **13**, e1005723 (2017).

- 754 24. B. Spitzer, L. Waschke, C. Summerfield, Selective overweighting of larger magnitudes
755 during noisy numerical comparison. *Nat. Hum. Behav.* **1** (2017).
- 756 25. D. Bernoulli, 1954 L. Sommer in *Econometrica* **22**, pp. 23-36., "Specimen Theoriae
757 Novae de Mensura Sortis" in *Comentarii Academiae Scientiarum Imperialis*
758 *Petropolitanae*, (1738), pp. 175–192.
- 759 26. R. D. Luce, B. A. Mellers, S. J. Chang, Is choice the correct primitive? On using certainty
760 equivalents and reference levels to predict choices among gambles. *J. Risk Uncertain.*
761 **6**, 115–143 (1993).
- 762 27. C. M. Constantinople, A. T. Piet, C. D. Brody, An Analysis of Decision under Risk in Rats.
763 *Curr. Biol.* **29**, 2066–2074.e5 (2019).
- 764 28. R. Gonzalez, G. Wu, On the Shape of the Probability Weighting Function. *Cogn.*
765 *Psychol.* **38**, 129–166 (1999).
- 766 29. D. Prelec, The Probability Weighting Function. *Econometrica* **66**, 497–527 (1998).
- 767 30. K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, K. J. Friston, Bayesian model
768 selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).
- 769 31. D. Kahneman, *Thinking, Fast and Slow* (Penguin, 2012).
- 770 32. R. McDermott, J. H. Fowler, O. Smirnov, On the Evolutionary Origin of Prospect Theory
771 Preferences. *J. Polit.* **70** (2008).
- 772 33. T. Caraco, Energy budgets, risk and foraging preferences in dark-eyed juncos (*Junco*
773 *hyemalis*). *Behav. Ecol. Sociobiol.* **8**, 213–217 (1981).
- 774 34. D. W. Stephens, The Logic of Risk-Sensitive Foraging Preferences. *Anim. Behav.* **29**,
775 628–629 (1981).
- 776 35. J. Apesteguia, M. A. Ballester, Monotone stochastic choice models: The case of risk and
777 time preferences. *J. Polit. Econ.* **126**, 74–106 (2018).
- 778 36. G. Loomes, R. Sugden, Testing different stochastic specifications of risky choice.
779 *Economica* **65**, 581–598 (1998).
- 780 37. J. Apesteguia, M. A. Ballester, Separating Predicted Randomness from Residual
781 Behavior. *J. Eur. Econ. Assoc.* **00**, 1–36 (2020).
- 782 38. R. Polanía, M. Woodford, C. Ruff, Efficient coding of subjective value. *Nat. Neurosci.*
783 **22**, 134–142 (2019).
- 784 39. J. A. Heng, M. Woodford, R. Polania, Efficient sampling and noisy decisions. *bioRxiv*,
785 799064 (2019).
- 786 40. K. Kobayashi, S. Ravaioli, A. Baranès, M. Woodford, J. Gottlieb, Diverse motives for
787 human curiosity. *Nat. Hum. Behav.* **3**, 587–595 (2019).
- 788 41. K. Kobayashi, M. Hsu, Neural mechanisms of updating under reducible and irreducible
789 uncertainty. *J. Neurosci.* **37**, 6972–6982 (2017).
- 790 42. P. Schwartenbeck, *et al.*, Computational mechanisms of curiosity and goal-directed
791 exploration. *Elife* **8**, 1–45 (2019).
- 792 43. R. C. Wilson, A. Geana, J. M. White, E. a Ludvig, J. D. Cohen, Humans Use Directed and
793 Random Exploration to Solve the Explore – Exploit Dilemma. *J. Exp. Psychol. Gen.* **143**,
794 2074–2081 (2014).
- 795 44. A. Clark, Whatever next? Predictive brains, situated agents, and the future of cognitive
796 science. *Behav. Brain Sci.* **36**, 181–204 (2013).
- 797 45. R. P. N. Rao, D. H. Ballard, Predictive coding in the visual cortex: a functional
798 interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87
799 (1999).
- 800 46. J. Hawkins, *On Intelligence: How a New Understanding of the Brain will Lead to the*

- Creation of Truly Intelligent Machines* (Times Books, 2004).
47. G. Pezzulo, P. Cisek, Navigating the Affordance Landscape : Feedback Control as a Process Model of Behavior and Cognition. *Trends Cogn. Sci.* **20**, 414–424 (2016).
 48. A. Tschantz, B. Millidge, A. K. Seth, C. L. Buckley, Reinforcement Learning through Active Inference. *arXiv*, 1–16 (2020).
 49. D. Gal, D. D. Rucker, The Loss of Loss Aversion: Will It Loom Larger Than Its Gain? *J. Consum. Psychol.* **28**, 497–516 (2018).
 50. E. Yechiam, Acceptable losses: the debatable origins of loss aversion. *Psychol. Res.* **83**, 1327–1339 (2019).
 51. C. Summerfield, K. Tsetsos, “Rationality and Efficiency in Human Decision Making” in *The Cognitive Neurosciences*, 6th Ed., D. Poeppel, G. R. Mangun, M. S. Gazzaniga, Eds. (MIT Press).
 52. E. Eldar, G. J. Bae, Z. Kurth-Nelson, P. Dayan, R. J. Dolan, Magnetoencephalography decoding reveals structural differences within integrative decision processes. *Nat. Hum. Behav.* **2**, 670–681 (2018).
 53. J. Daunizeau, V. Adam, L. Rigoux, VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Comput. Biol.* **10** (2014).

Supplementary Information

Encoding noise model

In previous work, the shape of the probability weighting function has been explained as a consequence of encoding noise, i.e. noise resulting from imprecise representation of the true value. Whilst our model assumes noise at a later stage (when prospects are compared), we nevertheless wanted to compare what predictions a model based on noise in probability encoding would make for our data. For this purpose, we implemented Steiner & Stewart's (1) model and compared its predictions on the lotteries taken from experiment 2 as these lotteries most closely resemble the scenario considered in their paper. Doing the comparison required some assumptions, however. First, we assumed that encoding of a lottery's probability is subject to Gaussian noise of magnitude σ . The Steiner & Stewart model is more general than this in only assuming non-degenerate noise distributions, but their predictions should nevertheless hold under the assumption of Gaussian noise. In order to approximate this noise, we took each of the m lotteries of the form (x_1, p_1) and expanded them into $n=100$ new "noisy" lotteries with (x_1, q_1) , where each new q_1 value represented a quantile of the normal distribution with mean p_1 and standard deviation σ . Quantiles were capped between $[1e-5, 1-1e-5]$, and each q_1 was enforced to fall within the interval $[0,1]$. Second, to allow direct comparison on the predictions for the value function, we assumed that values (including of the certain option) could also be distorted via x^κ , but no further noise was added to the value encoding step. The agent then made decisions between $N = m*n$ lotteries given parameters γ, κ, σ . The results of this comparison to our model (with and without entropy) are plotted in Fig S1. below. We note that these assumptions, while necessary for a comparison between models, were not part of their original implementation. Thus, the seemingly over-weighted values ($\log(\kappa) > 0$) may be remedied by further assumptions (e.g. a Bayesian scheme). However, this would substantially extend the original model.

References:

1. J. Steiner, C. Stewart, Perceiving Prospects Properly. *Am. Econ. Rev.* **106**, 1601–1631 (2016).

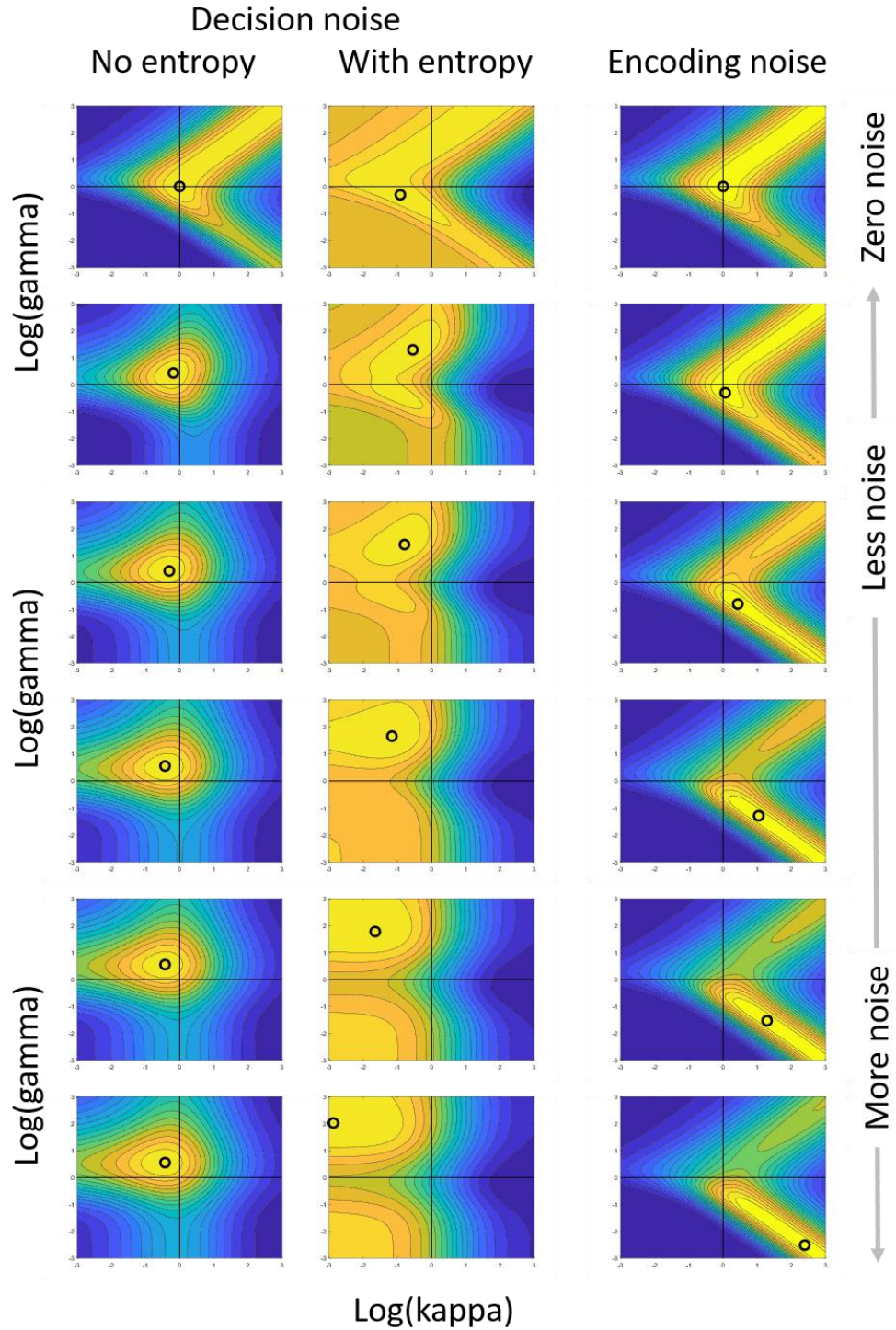


Figure S1. Comparison between loss landscapes of decision and encoding noise on the data from experiment 2. Each landscape shows the expected loss (brighter is better) for combinations of parameters kappa (each x-axis; governs the value function) and gamma (each y-axis; governs the shape of the Prospect Theory probability weighting function). Parameters are plotted in log space (base e), where the point (0,0) implies linear encoding of both probability and value (indicated by black lines). Parameters range from $[-3, 3]$. Black circles indicate optimal parameters given the noise. The first row shows the loss under zero noise, with noise then increasing with each row to the point where parameters change only little with further increases in noise. Left panel: Our model under the assumption of no entropic loss. Middle panel: our model with entropic noise assumption. Right panel: A model with encoding noise in probabilities (no noise in the value domain), based on (1).