nature human behaviour

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41562-019-0589-3

In the format provided by the authors and unedited.

# Diverse motives for human curiosity

**Kenji Kobayashi[1*], Silvio Ravaioli[2,3], Adrien Baranès[4], Michael Woodford[3] and Jacqueline Gottlieb[1,4,5]**

[1]The Mortimer B. Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA. [2]Sant'Anna School of Advanced Studies, Pisa, Italy. [3]Department of Economics, Columbia University, New York, NY, USA. [4]Department of Neuroscience, Columbia University, New York, NY, USA. [5]The Kavli Institute for Brain Science, Columbia University, New York, NY, USA. *e-mail: kenji.kobayashi@berkeley.edu

# Supplementary Notes

## Supplementary Note 1: Model derivation

Here, we derive the computational model that theoretically link our parameterization of choice behavior with the two motives for curiosity discussed in the text: reduction of uncertainty about future rewards and anticipatory utility.

The task that we consider is a choice between two information structures, which we denote $A$ and $B$. Each information structure specifies a set of possible information states, indexed by $i$, that can be reached when initial information is revealed, with *ex ante* probabilities $\pi_i$ at the time of choice. Each information state $i$ can be followed by a set of final states $(i, j)$, with probabilities $\pi_j$ conditional on state $i$ having been reached. In information state $i$, the outcome of lottery $x$ (the one revealed first) is revealed to be some particular value $x_i$, while lottery $y$ (the one yet to be revealed) can still have any of several possible outcomes $y_j$. In final state $(i, j)$, a reward $r_{ij} = x_i + y_j$ is obtained, the sum of the amounts earned from the two lotteries, with the value of $y_j$ being learned only when the final state is reached.

To model choices between two information structures, we consider that each potential information state $i$ is assigned a value $v_i$, which depends on the revealed outcome $x_i$ and the set of outcomes $y_j$ that remain possible. In the situation when an information structure is chosen, each information structure is assigned a value $U = f(v_1, v_2, \dots)$, which is an increasing function of the values of the information states that might be reached when the initial information is revealed. The probability of a decision maker's choosing $A$ over $B$ is then assumed to be an increasing function of the value difference $U^A - U^B$.

A familiar specification (expected utility theory) can be cast in this form, but does not allow a motive for curiosity. This kind of theory assigns a utility $u_{ij}$ to the outcomes in each of the possible final states $(i, j)$. It then values information states according to the expected utility to be obtained conditioning on information at that time,

$$v_i = \sum_j \pi_j \, u_{ij} \, , \tag{1}$$

and values the entire information structure according to the expected value $\bar{v}$ of the information states that can be reached,

$$U = \bar{v} \equiv \sum_i \pi_i v_i \, . \tag{2}$$

Equations (1) and (2) imply that $U$ depends only on the utilities of the possible final states, and the *ex-ante* probability of reaching each of them, as of the time of choice. The degree of interim information about which final outcomes are more likely to occur is irrelevant to the calculation of $U^A$ and $U^B$, and hence is irrelevant to choice.

We modify this specification in two respects, corresponding to the two possible motives for curiosity discussed in the main text. First, we follow Kreps and Porteus[1] and replace equation (2) by a nonlinear aggregator function. Specifically, we propose

$$U = \frac{1}{\alpha} \log \sum_i \pi_i \exp(\alpha v_i) \, . \tag{3}$$

This generalizes equation (2) in the sense that equation (3) approaches equation (2) in the limit as $\alpha \to 0$. If $\alpha > 0$, equation (3) implies that the aggregator function is strictly convex, so that Jensen's inequality implies that $U > \bar{v}$ whenever the $\{v_i\}$ are different from one another. Moreover, any mean-preserving spread of $\{v_i\}$ will increase the value of $U$ in this case, so that revelation of a greater amount of information at the interim stage (indicated by greater dispersion in the assessed values $\{v_i\}$ at that stage) will increase the value of that information structure. Thus, this specification incorporates a preference for early resolution of uncertainty when $\alpha > 0$, and the size of $\alpha$ measures the strength of that motive. (We can also allow $\alpha$ to be negative; in this case, the aggregator function is strictly concave, implying that $U < \bar{v}$ whenever the $\{v_i\}$ are different, so that the DM would prefer to have as little as possible early resolution of uncertainty.)

While this modification alone would allow one type of preference for non-instrumental information, it cannot account for any effect of ΔEV on participants' preferences between information structures. If the utility $u_{ij}$ assigned to any final outcome depends only on the total reward $r_{ij}$ received in that case, and the value assigned to an interim information state is given by some aggregator function of the final utilities that can be reached from it (as in the theory of Kreps and Porteus), then we can write $v_i$ as a function of the quantities $(r_{i1}, r_{i2}, \ldots)$. Suppose further that we write the possible outcomes of lottery $x$ as $x_i = \mu_x + \tilde{x}_i$, where $\mu_x$ is the mean outcome (the lottery's EV) and $\tilde{x}_i$ is the deviation from the mean outcome in the case of outcome $i$, and similarly with lottery $y$. Then we can write $v_i$ as a function of the quantity $\mu_x + \mu_y + \tilde{x}_i$ and the set of possible values $\{\tilde{y}_j\}$. It follows that the set of possible values $\{v_i\}$ is a function of the quantity $\mu_x + \mu_y$, the set of possible values $\{\tilde{x}_i\}$, and the set of possible values $\{\tilde{y}_j\}$. Therefore, the value $U$ must also be a function only of these quantities. This means that, while $U$ can depend on the sum of the EVs of the two lotteries $(\mu_x + \mu_y)$, it cannot depend on their difference ΔEV.

We accordingly generalize expected utility theory in a second respect as well. Following Caplin and Leahy[2], we allow the total utility obtained over a history $(i, j)$ to include not only on the utility obtained from receiving reward $r_{ij}$ eventually, but also a component of "anticipatory utility" from the interim information state $i$ itself. Specifically, in the particular computational model used in our work, we assume a total utility $r_{ij} + \phi x_i$, where the first term represents the consumption utility from actual receipt of an intrinsically valuable reward (assumed for simplicity to be linear in the size of the reward), and the second term the anticipatory utility in interim state $i$, in which one has already learned that one will receive the amount $x_i$. If $\phi > 0$, the anticipatory term corresponds to "savoring" of a future reward that one is already certain to receive, as proposed by Loewenstein[3], and the size of the coefficient $\phi$ measures the strength of this motive. (We can also allow for negative values of $\phi$; we assume only that $\phi > -1$, so that a larger lottery outcome $x_i$ still increases total utility along paths on which it occurs.) The value assigned to an interim information state $i$ is then given by

$$v_i = \phi x_i + \sum_j \pi_j \, r_{ij} \tag{4}$$

instead of equation (1). If $\phi > 0$, our theory incorporates a second potential reason for interest in obtaining information of no instrumental value, namely a desire to be able to savor rewards before actually receiving them.

An advantage of the particular functional forms in equations (3) and (4) is that, if we add a constant $w$ (possibly representing a fixed show-up fee for participation in the experiment, expected earnings from other trials of the experiment, or simply wealth that the participant will have from other sources) to the DM's wealth in each of the possible final outcomes $(i, j)$, so that $r_{ij} = w + x_i + y_j$, the value of $w$ has no effect on the utility difference $U^A - U^B$. (The addition of the constant amount $w$ simply increases $v_i$ by the amount $w$ for all interim states $i$ of both information structures, which in turn implies that the value of $U$ is increased by the amount $w$ for both information structures.) This means that our predictions about the frequency of choosing one information structure over another are independent of any assumptions about "narrow framing"; they are the same regardless of the degree to which the rewards obtained from the lotteries offered on a particular experimental trial are integrated with income that the participant can expect to have from other sources.

Finally, we complete our model with a particular quantitative specification of the way in which choice frequencies are determined by the utilities assigned to the different options. We assume that when a DM faces a choice between two information structures $A$ and $B$, the probability of choosing $A$ is given by

$$\mathrm{Prob}[A] = \frac{\exp(\lambda U^A)}{\exp(\lambda U^A) + \exp(\lambda U^B)}\ , \tag{5}$$

which, for some $\lambda > 0$, is an increasing (logistic) function of the utility difference $U^A - U^B$.

This familiar softmax specification (equation 5) can be given a variety of interpretations; it represents a version of the Luce[4] model of stochastic choice, and it can be interpreted as a random utility model[5] or a model of rational inattention[6]. An interpretation of particular interest to us is that it represents the choice frequencies implied by a drift-diffusion model (DDM)[7], in which the drift of the "decision value" is assumed to be proportional to the utility difference between the two options, as applied to binary choice experiments in the neuroeconomics literature[8]. This last interpretation has the advantage of allowing our model to also make predictions about how participants' response times vary with the properties of the two lotteries (**Fig. 4**; ***Methods***, equation (2)).

For the lotteries used in our experiment, the random variables $\tilde{x}$ and $\tilde{y}$ are always of a specific form, $\tilde{x} = \sigma_x \epsilon$ and $\tilde{y} = \sigma_y \epsilon$, where $\epsilon$ is a particular mean-zero random variable (taking five possible values, each with probability 1/5), and the factors $\sigma_x$ and $\sigma_y$ scale the variability of the lottery outcomes. Furthermore, the scale factors always take two particular values, $\sigma_{\mathrm{high}}$ and $\sigma_{\mathrm{low}}$, in the two lotteries used on a given trial. If the option $A$ reveals the high-variance lottery first, equations (3) and (4) imply that

$$U^A - U^B = \phi(\mu_x^A - \mu_x^B) + D(\alpha) \tag{6}$$

where

$$D(\alpha) \equiv \frac{1}{\alpha}\left[\log \sum\nolimits_i \pi_i \exp\big(\alpha(1+\phi)\sigma_{\mathrm{high}}\epsilon_i\big) - \log \sum\nolimits_i \pi_i \exp\big(\alpha(1+\phi)\sigma_{\mathrm{low}}\epsilon_i\big)\right]. \tag{7}$$

Moreover, because the random variable $\sigma_{\mathrm{high}}\epsilon$ is a mean-preserving spread of $\sigma_{\mathrm{low}}\epsilon$, Jensen's inequality implies that $D(\alpha)$ is positive, negative, or zero depending on whether $\alpha$ is positive, negative, or zero. Thus, the sign of $\alpha$ determines the sign of $D(\alpha)$, which is an increasing function of $\alpha$.

Substituting equation (6) into equation (5) yields the choice frequency prediction

$$\mathrm{Prob}[A] = \frac{1}{1 + \exp\big(-(\lambda D(\alpha) + \lambda\phi \cdot \Delta\mathrm{EV})\big)} \tag{8}$$

where $\Delta\mathrm{EV} = \mu_x^A - \mu_x^B$ is the difference in the EVs of the two lotteries.

This equation (8) has the logit form used in the main text (***Methods***, equation (1)). Note that our model predicts that for a given $\Delta$EV, when the option A reveals the high-variance lottery first, the probability of choosing $A$ will be greater (smaller) if $\alpha$ (and $w_{\mathrm{var}}$) is positive (negative). At the same time, it predicts that for given variances of the two lotteries, when the option A reveals the high-EV lottery first, the probability of choosing $A$ will be greater (smaller) if $\phi$ (and $w_{\Delta\mathrm{EV}}$) is positive (negative). Thus, the parameters $\alpha$ and $\phi$ respectively determine the effects of differential variance and ΔEV on the lottery that a DM is predicted to prefer to learn about earlier.

Because there are only two coefficients that can be estimated from behavior— the intercept ($w_{\mathrm{var}}$) and the coefficient of sensitivity to ΔEV ($w_{\Delta\mathrm{EV}}$) — we cannot separately estimate the values of the three parameters $\lambda$, $\alpha$, and $\phi$ of our model. However, estimation of the two coefficients allows us to determine the respective signs and magnitudes of $\alpha$ and $\phi$, i.e., the relative strength of the two motives for curiosity for each of our participants.

## Supplementary Note 2: Generalization to allow anticipatory utility from rewards that are not certain to be received

In the model proposed in **Supplementary Note 1**, we assumed that an additional contribution to utility from the anticipation of a reward that one has already learned that one will receive occurs only in the case that one is certain to receive the reward. For this reason, equation (4) included a contribution to utility $\phi x_i$ from the anticipation of the reward that is already certain in interim state $i$, but no similar contribution from any of the possible rewards from the other lottery $\{y_j\}$, each of which will be received only with probability 1/5.

The concept of anticipatory utility, however, does not require this to be the case. One might equally logically suppose that there is also savoring of the possibility of receiving a reward even when it is not yet certain that one will get it. (The original discussion of savoring of future rewards by Loewenstein[3] does not take a stand on this issue, as it deals only with situations in which all future rewards are completely certain.) Indeed, it makes sense to suppose that rewards that are not totally certain can be savored to some extent, given that few outcomes can be regarded as totally certain in life.

We can generalize our model to allow savoring of both certain future rewards and ones that are only possibilities. We allow them potentially to be savored to differing extents; the generalized model nests the case in which there would be no difference in savoring of rewards of the two types, but also allows us to suppose that more certain outcomes may be savored to a greater extent. (We can remain agnostic on this issue *a priori* and allow our empirical data to resolve it.)

In this more general specification, the value assigned to interim information state $i$ will be of the form

$$v_i = \varphi(1)x_i + \sum_j \varphi(\pi_j)y_j + \sum_j \pi_j r_{ij} \tag{9}$$

instead of equation (4), where $\varphi(\pi)$ indicates the degree to which an outcome with probability $\pi$ is savored (conditional on being in the information state in which the savoring occurs). The first term on the right-hand side indicates the utility from savoring the known outcome $x_i$ from the lottery revealed first (this outcome has probability 1, once information state $i$ is reached); the second term sums the contributions to utility from savoring each of the outcomes $y_i$ from the lottery that is yet to be revealed (which can still occur with probability $\pi_j$); and the final term indicates the direct utility from eventual receipt of the lottery payouts, as in equation (4).

We shall assume in general that $\varphi(\pi)$ is a non-decreasing function (an outcome that is more certain to occur cannot be savored less because of that) with $\varphi(0) = 0$ (an outcome that is revealed to be impossible is not savored). We also assume that $\varphi(\pi) > -\pi$ for all $\pi$, so that an increase in any of the lottery outcomes results in the higher value when both direct and anticipatory effects on utility are considered.

In the case of the lotteries used in our experiment (each lottery has 5 equally probable outcomes), equation (9) can be written as

$$v_i = \psi \tilde{v}_i \,, \tag{10}$$

where

$$\psi \equiv 1 + 5\varphi(0.2) > 0$$

and $\tilde{v}_i$ refers to the old definition of $v_i$ given in equation (4), setting

$$\phi \equiv \frac{\varphi(1) - 5\varphi(0.2)}{1 + 5\varphi(0.2)} \,.$$

Because the value of each interim state is given by the same formula as before, except for multiplication by the positive factor $\psi$, the predictions of the model are the same as before, except with the parameter $\alpha$ replaced by $\alpha\psi$ and the parameter $\lambda$ replaced by $\lambda\psi$. In particular, equation (8) now becomes

$$\text{Prob}[A] = \frac{1}{1 + \exp\big(-(\lambda\psi D(\alpha\psi) + \lambda\psi\phi \cdot \Delta\text{EV})\big)}.$$

Thus we again conclude, in this more general model, that the coefficient multiplying ΔEV in our empirical model can be interpreted as a positive multiple of $\phi$, so that the sign of this coefficient tells us the sign of $\phi$. The interpretation of the parameter $\phi$, however, is now somewhat different; in this generalized model, it measures not the strength of anticipatory utility from the savoring of future rewards in general, but rather the degree to which the anticipatory utility from more certain future rewards is greater than that resulting from the anticipation of less probable future rewards, to an extent that is more than proportional to the increase in the rewards' probability. That is, a finding that $w_{\Delta\text{EV}} > 0$ implies that $\phi > 0$, which in the generalized model implies that

$$\varphi(1) > 5\varphi(0.2) \,. \tag{11}$$

If $\varphi(\pi)$ is a linear function of $\pi$ (that is, the degree to which an outcome is savored in advance is proportional to the probability of its occurrence), then $\varphi(1) = 5\varphi(0.2)$ and equation (11) does not hold, implying $\phi = 0$. But if $\varphi(\pi)$ is an increasing, strictly convex function (for example, if it is a differentiable function with derivatives $\varphi' > 0, \varphi'' > 0$), then $\varphi(\pi)/\pi$ is an increasing function of $\pi$, which implies that equation (11) holds and $\phi > 0$. In this case, the model predicts the behavior of the large subset of our participants who exhibited significantly positive $w_{\Delta\text{EV}}$. Note that we do not assume that such behavior indicates that savoring of a certain outcome is materially different from savoring of an outcome with a very high probability; our interpretation is perfectly consistent with the (very plausible) assumption that $\varphi(\pi)$ is a continuous function of $\pi$, and hence that the valuations of the different options should be continuous functions of the probabilities defining the two lotteries.

## Supplementary Note 3: Comparison with the model of Iigaya et al. (2016)

Certain aspects of our model resemble a related model proposed by Iigaya et al. (2016)[9]. Here we discuss the similarities and differences between the two models in more detail.

Like us, Iigaya et al. propose a model of information choice with a recursive structure: the frequency of choice between two options $A$ and $B$ is a function of the difference $U^A - U^B$ between the values assigned to those two options; the value $U^A$ assigned to an option $A$ is an increasing function of the values $v_i$ assigned to the different possible interim information states $i$ that can be reached if that option is chosen; and the value $v_i$ assigned to an interim information state $i$ is an increasing function of the values of the rewards received in the various final states $(i, j)$ that can be reached from that interim state. In Iigaya et al's case, this recursive structure follows from a model of reinforcement learning, in which the values that are learned for cues depend on the values that are learned for the states that succeed them. But whether the recursive structure of valuations is assumed to be acquired through a reinforcement learning mechanism has no consequences for the predictions of interest in their paper or for the interpretation that their model would provide of our results; all that matters is the assumed recursive structure, which we also assume.

The model of Iigaya et al. is also similar to ours insofar as a key element of their model is the assumption that one of the contributions to the value $v_i$ comes from anticipatory savoring of the rewards to be received in states that can follow the interim state $i$. Like us, they assume that the value $v_i$ can be written as the sum of two components, one representing the utility from savoring and one representing the direct utility from rewards once they are obtained. (See equation (1) of their paper.) In their discussion, one of the main reasons to distinguish the two components is that they are predicted to be affected differently by changes in the length of the time delay between the informational cue and the eventual receipt of the rewards; we do not discuss such effects, since our experiment did not vary the length of the delay. In our discussion, instead, the crucial difference between the components is that they depend differently on the probabilities of the different possible outcomes, which Iigaya et al. do not discuss.

The model of Iigaya et al. differs from ours in two important respects. First, they propose that the contribution to the value of informational state $i$ that comes from savoring depends not only on the rewards that can follow that state and their probabilities, but also on the size of the reward prediction error (RPE) associated with that state. Specifically, they assume that the value that is learned for an informational state $i$ is of the form

$$v_i = \eta_i s_i + R_i \,. \tag{12}$$

In this formula, $R_i \equiv \sum_j r_{ij}$ is the expected (discounted) utility from eventually receiving the rewards that follow $i$. The quantity $s_i$ indicates the total amount of savoring following receipt of cue $i$, prior to "boosting" — assumed, as in our model, to be a function of the rewards that can be received following cue $i$ and their probabilities. The multiplicative factor $\eta_i$ represents the degree to which the savoring consequent upon receiving cue $i$ is "boosted" by the attention drawn to cue $i$. The boosting factor $\eta_i$ is assumed to be an increasing function of $|\delta_i|$, where $\delta_i = v_i - \overline{v}$ is the RPE associated with receipt of cue $i$, and $\overline{v} \equiv \sum_i \pi_i v_i$ is the expected value of the cue, after the option has been chosen but before receipt of the cue. The value $U^A$ assigned to an option $A$ is then simply the expected value $\overline{v}$ of the informational states that may be reached following that choice.

The second important difference in the model of Iigaya et al. is that the different possible rewards following a cue contribute to the utility from savoring in exactly the same proportion as they contribute to the expected reward $R_i$. Using the notation introduced in **Supplementary Note 2**, they assume that $\varphi(\pi) = s \cdot \pi$ for some constant $s > 0$, so that $s_i = s \cdot R_i$. This implies that their model does not satisfy equation (11).

Equation (12) and the functional relation between $\delta_i$ and $\eta_i$ suffice to determine the values $v_i$ of each of the interim informational states, using as inputs only the probabilities $\pi_i$ and the values $R_i$ for each of the states. Since in our experiment, the probabilities $\{\pi_i\}$ are always the same, and $R_i = \mu_x + \mu_y + \sigma_x \epsilon_i$, where the

$\{\epsilon_i\}$ are the same in all cases as well, this model would predict that the only parameters that can have any effect on the values $\{v_i\}$ are the sum $\mu_x + \mu_y$ and $\sigma_x$, the parameter determining the variability of the lottery that is revealed first. Furthermore, since $\mu_x + \mu_y$ was the same for both choices on a given trial, this parameter cannot bias choice toward either of the options. Thus, according to Iigaya et al.'s model, the only possible reason for a preference for early revelation of one lottery over the other is the difference in the degree of variability of the outcomes of the two lotteries, just as in Kreps and Porteus's model discussed in **Supplementary Note 1**.

Thus, Iigaya et al.'s model does provide an alternative to the Kreps-Porteus hypothesis of nonlinear aggregation of the values of the possible information states, as an explanation for why some participants preferred to have high-variance lotteries revealed earlier (i.e., $w_{\text{var}} > 0$). (It is not entirely unrelated to the Kreps-Porteus explanation, however; the Iigaya et al. model of attentional "boosting" is essentially a form of nonlinear aggregation of the values associated with the different interim states.) However, their model provides no explanation for our empirical finding that $w_{\Delta\text{EV}}$ was significantly different from zero for many of our participants.

## Supplementary Note 4: Unitary vs mixed strategy

A good number of our participants exhibited both effects of ΔEV and variance. This might have been caused because of a mixture of two types of behavior, one reward-seeking and one information-seeking. Although we show that overall behavior was clearly sensitive to the task (**Supplementary Figures 2, 3**) and systematic misunderstanding of the task is quite unlikely to explain our results, it might be possible that some participants fell back to reward-maximizing behavior in some trials, as if they had been involved in the risk-taking task. This scenario is worth considering because, if this was the case, the effect of ΔEV could be entirely attributed to the erroneous reward-maximizing choices.

We conducted additional analysis to examine such a mixed strategy. We compared the standard model (***Methods***, equation (1)), which assumes a unitary strategy of information-seeking characterized by both ΔEV and variance effects, with an alternative model describing a mixed-strategy, whereby people made information-seeking choices on some trials based on variance only and lapsed into (erroneous) reward-maximizing choices on others. Under this mixed-strategy model, the choice probability is described as

$$\text{Prob[high-variance choice]} = \frac{q}{1 + \exp(-(w_0 + w_{\Delta\text{EV}} \cdot \Delta\text{EV}))} + \frac{1 - q}{1 + \exp(-w_{\text{var}})} \qquad (13)$$

in which the first term on the right-hand side corresponds to reward-maximizing choices and the second term to information-seeking choices based only on variance. This model has four free parameters: $q$ is the probability of reward-maximizing choices ($0 < q < 1$) (i.e., $(1 - q)$ is the probability of information seeking choices), $w_0$ and $w_{\Delta\text{EV}}$ measure the individual's risk attitude and sensitivity to EV in reward-maximizing choices (applying ***Methods*** equation (1)), and $w_{\text{var}}$ measures the uncertainty effect in information-seeking choices (applying ***Methods*** equation (1) and set $w_{\Delta\text{EV}}$ to 0).

We compared these two models at the individual level without regularization. We concluded that a participant showed exhibited the mixed strategy if three conditions were met: *a*) they showed evidence of reward-maximizing choices in the original analysis ($w_{\Delta\text{EV}} > 0$, $p < 0.05$), *b*) the sign of parameters $w_0$ estimated from the risk-taking task using the standard model (***Methods***, equation (1)) was consistent with the sign of $w_0$ estimated from the observing task using equation (13), indicating the consistency in risk attitudes (recall that in the risk-taking task, positive $w_0$ indicates risk loving and negative $w_0$ indicates risk aversion), and $c$) the mixed-strategy model fitted the observing task behavior better than by the standard model. For the condition $c$, we used likelihood ratio test ($\alpha = 0.05$), because the standard model is nested in the mixed-strategy model.

These three conditions were met only by 14 participants in the Gain domain and 7 participants in the loss domain. Importantly, no participant met all conditions in both domains. After removing these participants, the population statistics remained effectively the same (**Supplementary Table 1**). We can thus conclude our results cannot be explained by this sort of mixed strategies.

## Supplementary References


1 Kreps, D. M. & Porteus, E. L. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica* **46**, 185-200 (1978).

2 Caplin, A. & Leahy, J. Psychological expected utility theory and anticipatory feelings. *Q. J. Econ.* **116**, 55-79 (2001).

3 Loewenstein, G. Anticipation and the valuation of delayed consumption. *Econ. J.* **97**, 666-684 (1987).

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

5 McFadden, D. L. Structural discrete probability models derived from theories of choice. in *Structural Analysis of Discrete Data and Econometric Applications* (ed C. F. Manski and D. L. McFadden) (MIT Press, 1981).

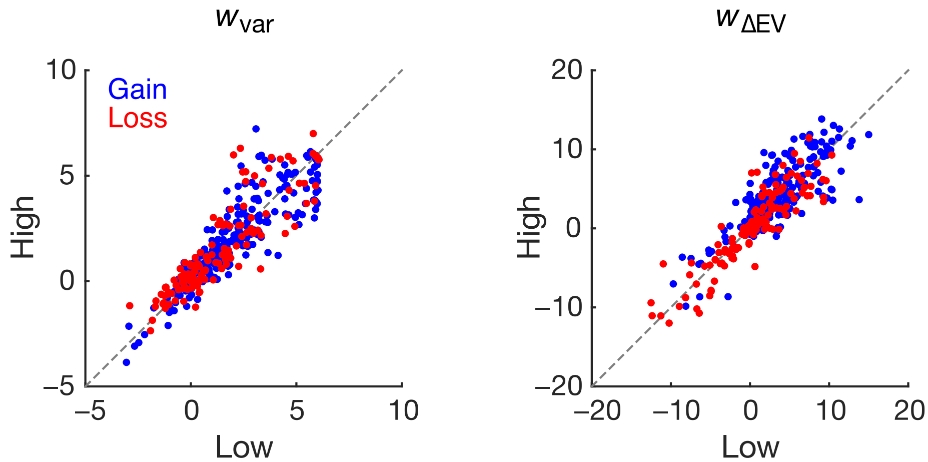
6 Matejka, F. & McKay, A. Rational inattention to discrete choices: a new foundation for the multinomial logit model. *Am. Econ. Rev.* **105**, 272-298 (2015).

7 Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R. & Yang, T. The speed and accuracy of a simple perceptual decision: A mathematical primer. in *Bayesian Brain* (ed K. Doya et al.) (MIT Press, 2007).

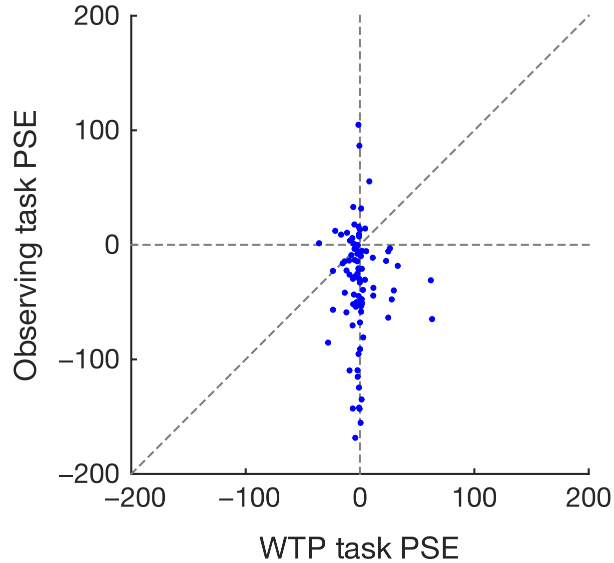
8 Fehr, E. & Rangel, A. Neuroeconomic foundations of economic choice – recent advances. *J. Econ. Perspect.* **25**, 3-30 (2011).

9 Iigaya, K., Story, G. W., Kurth-Nelson, Z., Dolan, R. J. & Dayan, P. The modulation of savouring by prediction error and its effects on choice. *eLife* **5**, e13747 (2016).

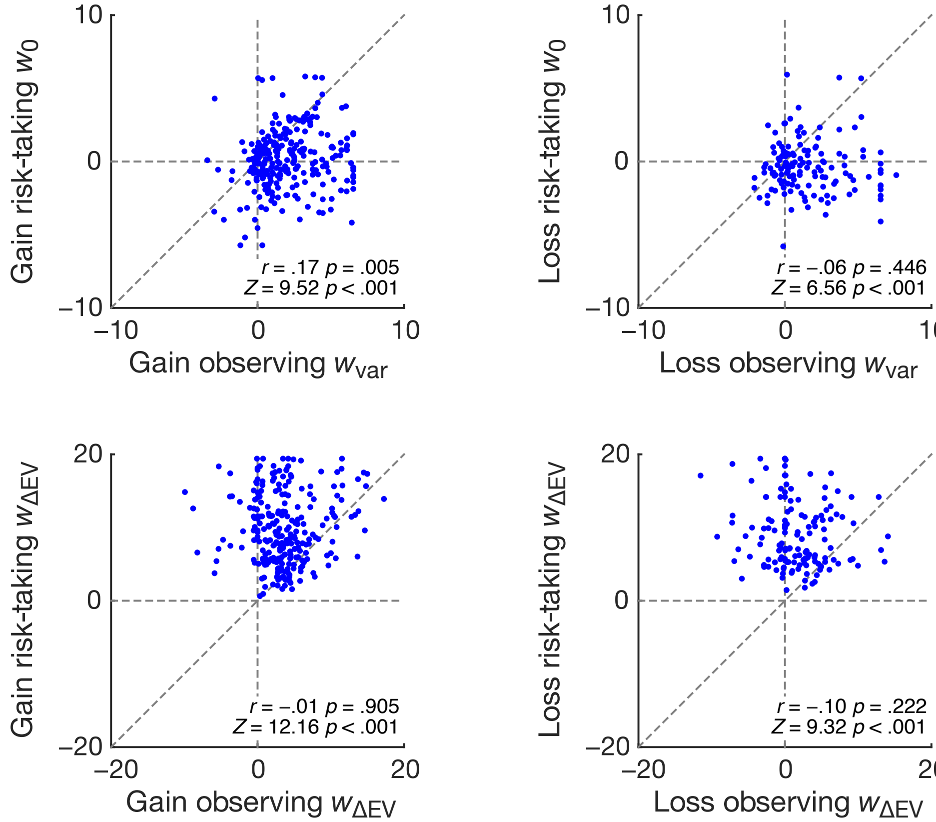
## Supplementary Figures



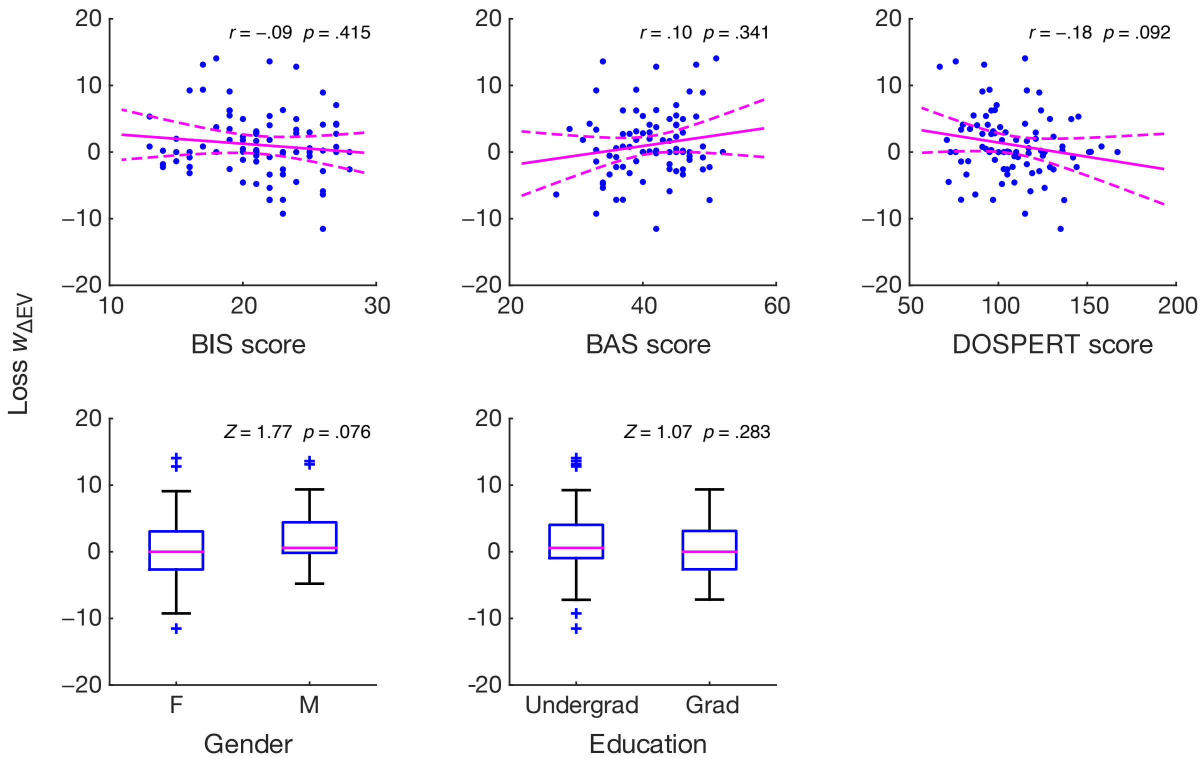

**Supplementary Figure 1. Parameter estimates do not depend on the absolute level of EV.** Comparisons of $w_{var}$ and $w_{\Delta EV}$ for trials in which the high-variance lottery's EV was higher or lower than the median (across all trials) for each participant. Each point represents one participant tested in the gain (blue) or loss (red) domains. All the coefficients are highly correlated for both gains ($w_{var}$: Spearman rho = 0.92, $w_{\Delta EV}$: rho = 0.79, both $p < .001$, $n$ = 257) and losses ($w_{var}$: rho = 0.91, $w_{\Delta EV}$: rho = 0.90, both $p < .001$, $n$ = 140).




**Supplementary Figure 2. PSE values were much smaller (less negative) on the WTP paradigm relative to the 2-lottery observing paradigm.** This plot shows the 90 participants who completed both paradigms and showed sensitivity to $w_{\Delta EV}$ in both paradigms. Participants showed a significant willingness to pay for information on the WTP paradigm (PSE median = −1.26, mean = 0.08, SEM = 1.55, $Z$ = 2.17, $p$ = .030). At the individual level, 42 participants showed a significant WTP, of whom the majority (31) paid for information (i.e., negative PSE, bootstrap $p < 0.05$) while 11 participants paid to avoid information (positive PSE, $p < 0.05$). However, PSEs in the WTP task were much smaller (less negative) than in the 2-lottery observing paradigm ($Z$ = 5.53, $p < 0.001$; the 2-lottery paradigm PSE median = −26.8, mean = −33.9, SEM = 5.19).




**Supplementary Figure 3. Compared to risk taking, behavior in the observing task is more driven by variance and less by EV.** In each plot, individual parameter estimates are compared between the tasks, in Gain (left) and Loss (right) domain. Numbers indicate Spearman correlation ("*r*"), Wilcoxon signed rank test statistics ("*Z*"), and their *p* values. $w_0$ (referred to as $w_{var}$ in the observing task; see ***Methods***) was larger in the observing task than the risk-taking task (*top*), while $w_{\Delta EV}$ was smaller in the observing task than the risk-taking task (*bottom*). These confirm that participants were sensitive to the task instruction and the incentive structure. Inter-task correlation was found only for $w_{var}$ in Gain domain (*upper left*), indicating that risk-seeking was weakly correlated with the tendency to reduce uncertainty in information seeking. This association was not observed in Loss domain or with $w_{\Delta EV}$, suggesting that information-seeking motivations are overall distinct from behaviorally measured risk attitude (see also personality regression results in **Supplementary Figure 4** and **Supplementary Tables 2-4**).




**Supplementary Figure 4. Association between $w_{\Delta EV}$ in Loss domain and variables selected in Lasso regression.** Shown in each plot is linear regression ("*r*") or Wilcoxon rank sum test ("*Z*") and its *p* value, evaluated in isolation, i.e., without any other predictors included in modeling. None of the statistical tests is significant, indicating weak association between the observing behavior and personality / demographic measures. *Top row*, the solid line: linear regression fit, dashed line: 95% CI. *Bottom row*, center line: median, box limits: upper and lower quartiles, whiskers: 1.5 × interquartile range, points: outliers.

## Supplementary Tables

**Supplementary Table 1. Population statistics across participant groups.** Shown are the summary statistics of parameter estimates in Gain and Loss domains and their Gain-Loss correlation (among participants tested in both domains). For Gain and Loss domains, the first number in each cell is median, the second number is $Z$ statistics of Wilcoxon signed rank test (in parenthesis), and the third number is its $p$ value. For Gain-Loss correlation, the first number in each cell is Spearman correlation and the second number is its $p$ value. $P$ values smaller than 0.05 are shown in bold. The bottom four rows show the results from analyses that excluded participants who may have used a mixed strategy (**Supplementary Note 4**) and those that included the participants that were excluded from the main analyses because they did not respond to incentives on the risk-taking or WTP tasks (see ***Methods***). G: Gain domain, L: Loss domain.

| Group | | $n$ | $w_{var}$ | | | $w_{\Delta EV}$ | | | Task order |
|---|---|---|---|---|---|---|---|---|---|
| | | | Gain | Loss | G-L corr. | Gain | Loss | G-L corr. | |
| **Lab I (individual)** | | 39 | 2.32<br>(4.33)<br>**$1.5\cdot10^{-5}$** | | | 2.97<br>(4.51)<br>**$6.6\cdot10^{-6}$** | | | risk(G) WTP(G)<br>observing(G) |
| **AT** | | 132 | 1.39<br>(8.65)<br>**$5.0\cdot10^{-18}$** | | | 3.46<br>(9.51)<br>**$2.0\cdot10^{-21}$** | | | |
| **G only** | | 78 | 1.48<br>(6.60)<br>**$4.2\cdot10^{-11}$** | | | 3.48<br>(7.39)<br>**$1.4\cdot10^{-13}$** | | | risk(G) WTP(G)<br>observing(G) |
| **G first, L second** | | 54 | 1.35<br>(5.64)<br>**$1.8\cdot10^{-8}$** | 0.14<br>(2.42)<br>**.015** | 0.55<br>**$2.0\cdot10^{-5}$** | 3.41<br>(5.95)<br>**$2.6\cdot10^{-9}$** | 2.57<br>(4.63)<br>**$3.7\cdot10^{-6}$** | 0.50<br>**$1.3\cdot10^{-4}$** | risk(G) observing(G)<br>risk(L) observing(L) |
| **Lab II (group)** | | 86 | 1.60<br>(7.14)<br>**$9.5\cdot10^{-13}$** | 1.45<br>(6.33)<br>**$2.4\cdot10^{-10}$** | 0.71<br>**$2.2\cdot10^{-14}$** | 2.98<br>(5.20)<br>**$2.0\cdot10^{-7}$** | 0.21<br>(1.90)<br>.057 | 0.43<br>**$3.3\cdot10^{-5}$** | |
| **G first, L second** | | 43 | 1.53<br>(4.67)<br>**$3.0\cdot10^{-6}$** | 1.62<br>(4.70)<br>**$2.6\cdot10^{-6}$** | 0.66<br>**$1.5\cdot10^{-6}$** | 3.33<br>(4.08)<br>**$4.5\cdot10^{-5}$** | 0.78<br>(2.23)<br>**.026** | 0.71<br>**$1.1\cdot10^{-7}$** | observing(G)<br>observing(L)<br>risk(G) risk(L) |
| **L first, G second** | | 43 | 1.71<br>(5.39)<br>**$7.2\cdot10^{-8}$** | 1.07<br>(4.21)<br>**$2.5\cdot10^{-5}$** | 0.77<br>**$1.3\cdot10^{-9}$** | 2.64<br>(3.20)<br>**$1.4\cdot10^{-3}$** | $1.7\cdot10^{-7}$<br>(0.48)<br>.629 | 0.13<br>.411 | observing(L)<br>observing(G)<br>risk(L) risk(G) |
| **Entire sample** | | 257 | 1.53<br>(11.89)<br>**$1.3\cdot10^{-32}$** | | | 3.32<br>(11.55)<br>**$7.6\cdot10^{-31}$** | | | |
| **Completed G & L** | | 140 | 1.49<br>(9.09)<br>**$9.7\cdot10^{-20}$** | 0.90<br>(6.77)<br>**$1.3\cdot10^{-11}$** | 0.65<br>**$5.0\cdot10^{-18}$** | 3.31<br>(7.73)<br>**$1.1\cdot10^{-14}$** | 0.96<br>(4.30)<br>**$1.7\cdot10^{-5}$** | 0.49<br>**$8.5\cdot10^{-10}$** | |
| **Mixed-strategy participants excluded** | **G** | 243 | 1.54<br>(11.56)<br>**$6.9\cdot10^{-31}$** | | | 3.30<br>(11.09)<br>**$1.4\cdot10^{-28}$** | | | |
| | **L** | 133 | | 0.96<br>(6.77)<br>**$1.3\cdot10^{-11}$** | | | 0.78<br>(3.76)<br>**$1.7\cdot10^{-4}$** | | |
| **Non-monetarily-incentivized participants included** | **G** | 298 | 1.29<br>(12.17)<br>**$4.6\cdot10^{-34}$** | | | 3.05<br>(12.58)<br>**$2.1\cdot10^{-36}$** | | | |
| | **L** | 159 | | 0.56<br>(6.39)<br>**$1.7\cdot10^{-10}$** | | | 0.83<br>(4.77)<br>**$1.8\cdot10^{-6}$** | | |

**Supplementary Table 2. Variable selection results of personality regression on $w_{\Delta EV}$ in Loss domain.** Only predictors selected by the automated Lasso procedure are shown (see ***Methods***). Note that the statistics shown here were obtained without regularization and presented just for reference.

| Predictor | Coefficients | SE | *t(80)* | *p* |
|---|---|---|---|---|
| BIS | −0.151 | 0.108 | −1.404 | 0.164 |
| BAS | 0.272 | 0.105 | 2.583 | 0.012 |
| DOSPERT | −0.452 | 0.115 | −3.912 | < 0.001 |
| gender | −0.648 | 0.216 | −2.998 | 0.004 |
| education | −0.267 | 0.200 | −1.332 | 0.188 |

**Supplementary Table 3. The effect of task order in personality regression on $w_{\Delta EV}$ in Loss domain.** In this linear regression (without regularization), interaction between personality measures selected by the variable selection procedure (**Supplementary Table 2**) and the task order (gain / loss) were tested. None of the interaction terms was not significant, confirming that the relationship between behavior and personality measures was not confounded by the task order.

| Predictor | Coefficients | SE | *t(76)* | *p* |
|---|---|---|---|---|
| BIS | −0.179 | 0.109 | −1.642 | 0.105 |
| BAS | 0.221 | 0.109 | 2.021 | 0.047 |
| DOSPERT | −0.448 | 0.115 | −3.882 | < 0.001 |
| BIS × order | −0.831 | 0.596 | −1.394 | 0.167 |
| BAS × order | −1.009 | 0.834 | −1.209 | 0.230 |
| DOSPERT × order | 0.060 | 0.576 | 0.105 | 0.917 |
| gender | −0.643 | 0.225 | −2.862 | 0.005 |
| education | −0.330 | 0.207 | −1.597 | 0.114 |
| order | 3.711 | 2.298 | 1.615 | 0.110 |

**Supplementary Table 4. Personality regression on $w_{\Delta EV}$ in Loss domain using scale sub-measures.** Two measures selected by the original Lasso regression (**Supplementary Table 2**), BAS and DOSPERT, were replaced with their sub-measures in linear regression (without regularization). The results suggest that BAS reward responsiveness (e.g., "when good things happen to me, it affects me strongly"), DOSPERT ethical (e.g., questionable tax deductions), and DOSPERT social (e.g., disagreement with authority) sub-measures are relatively strongly associated with $w_{\Delta EV}$ in Loss domain.

| Predictor | Coefficients | SE | *t(74)* | *p* |
|---|---|---|---|---|
| BIS | −0.255 | 0.112 | −2.277 | 0.026 |
| BAS drive | 0.045 | 0.108 | 0.420 | 0.676 |
| BAS fun seeking | 0.108 | 0.120 | 0.906 | 0.368 |
| BAS reward responsiveness | 0.193 | 0.112 | 1.718 | 0.090 |
| DOSPERT ethical | −0.245 | 0.105 | −2.333 | 0.022 |
| DOSPERT financial | 0.035 | 0.107 | 0.329 | 0.743 |
| DOSPERT health / safety | 0.025 | 0.127 | 0.194 | 0.847 |
| DOSPERT recreational | −0.207 | 0.116 | −1.782 | 0.079 |
| DOSPERT social | −0.360 | 0.109 | −3.299 | 0.001 |
| gender | −0.459 | 0.215 | −2.138 | 0.036 |
| education | −0.232 | 0.200 | −1.157 | 0.251 |