# Efficient Coding and Risky Choice

Cary Frydman[†] and Lawrence J. Jin[§]

This Version: October, 2018

**Abstract:** We present a model of risky choice in which the perception of a lottery payoff is noisy and optimally depends on the payoff distribution to which the decision maker has adapted. The perceived value of a payoff is precisely defined according to a core idea in neuroscience called the *efficient coding hypothesis*, which indicates that more perceptual resources are allocated to those stimuli that occur more frequently. We show that this principle implies that, for a given choice set of lotteries, risk taking varies systematically with the recently encountered distribution of payoffs. We test our model in two laboratory experiments. In the first experiment, we manipulate the distribution from which payoffs are drawn. Consistent with efficient coding of lottery payoffs, we find that risk taking is more sensitive to payoffs that are encountered more frequently in the choice set. Furthermore, sensitivity to extreme payoffs is initially small, but grows over time after repeated exposure. Our second experiment consists of a purely perceptual task, in which subjects classify which of two symbolic numbers is larger. We find that accuracy depends on the distribution of numbers to which the subject has adapted, which provides support for our key model assumption that perception of numerical payoffs is noisy and changes across environments.

[*] We thank Nicholas Barberis, Elise Payzan-LeNestour, Antonio Rangel, Leifu Zhang, and seminar participants at the Chinese University of Hong Kong, the University of New South Wales, the University of Technology Sydney, the University of Utah, Washington University, and the Sloan-Nomis Workshop on the Cognitive Foundations of Economic Behavior for helpful comments. Frydman thanks the NSF for financial support.

[‡] Marshall School of Business at USC cfrydman@marshall.usc.edu, +1 (213) 821-5586.

[§] Division of the Humanities and Social Sciences at Caltech. lawrence.jin@caltech.edu, +1 (626) 395-4558.

## I. Introduction

When choosing between two lotteries, the decision maker (called "*DM*" hereafter) first perceives the set of payoffs from each lottery, and then executes a decision. Because there are constraints on the degree to which the brain can process information, the perception of numerical payoffs is inherently imperfect (Dehaene, 2011). Understanding precisely how these constraints affect perception has the potential to generate new insights about risk taking, and in particular, its instability over time. For example, decades of experiments have shown that one source of instability is the sequence of past outcomes that the subject experiences: past gains and losses have a systematic effect on subsequent risk taking (Thaler and Johnson, 1990; Weber and Camerer, 1998; Imas, 2016). A different potential source of instability is variation in perception, which can occur independently of past outcomes or changes in wealth.

Why would the *DM*'s perception of a given risky payoff vary across different environments? If the mechanism used for perceiving payoffs is similar to the one used for perceiving sensory stimuli such as light or sound, then it may in fact be optimal to hold different perceptions of the same payoff in different environments. Specifically, a core idea in neuroscience called the *efficient coding hypothesis* states that the brain should allocate resources so that perception is more sensitive to those stimuli that are expected to occur more frequently (Barlow, 1961; Laughlin, 1981). This principle explains why we are temporarily "blinded" when moving from a dark room to a brightly lit one, because resources have not yet been adjusted for perceiving objects in the new bright environment. If the efficient coding hypothesis extends to the domain of risky choice, this can provide a normative explanation for the systematic variation of risk taking across environments.

In this paper, we present a model of choice under risk in which the perception of payoffs is governed by efficient coding; we then test the model experimentally to assess whether risk taking varies with the recently encountered payoff distribution. Our model builds on the recent theoretical work of Woodford (2012) and Khaw, Li, and Woodford (2018) (hereafter "KLW"), who assume that the perception of risky payoffs is imperfect and is estimated through Bayesian inference. Our main theoretical contribution is to provide, within the KLW framework, a

microfoundation for imperfect perception of risky payoffs that is governed by efficient coding. Specifically, we derive the optimal set of likelihood functions according to a capacity constraint and the payoff distribution to which the *DM* has adapted. This jointly constrains the *DM*'s prior and the likelihood functions through the distribution of payoffs, which provides an extra layer of discipline in the Bayesian framework (Wei and Stocker, 2015; 2017). We then show theoretically that, for a given choice set, risk taking will vary systematically with the payoff distribution to which the subject has adapted. Our main theoretical prediction is that both perception and risk taking are more sensitive to payoffs that are more likely to appear in the *DM*'s choice set.

An important consequence of this result is that the subjective value function is malleable, and undergoes precisely defined changes when there is a shift in the payoff distribution. For example, if the upside of a risky lottery is often in the range between $10 and $20, then perceptual resources are allocated towards discriminating between payoffs in this range. In this same environment, if the upside is occasionally increased from $30 to $40, then risk taking will not increase much because the *DM*'s perceptual system cannot easily distinguish between these two infrequent amounts. However, if the overall distribution of payoffs changes, so that the upside frequently falls between $30 and $40, then the *DM* perceives this difference to be large, and risk taking will increase substantially when the upside is increased from $30 to $40. Thus, diminishing sensitivity emerges as part of the optimal solution of allocating resources away from payoffs that are unlikely to appear in the choice set. As in Woodford (2012) and KLW, our model predicts that the value function is itself stochastic, but our efficient coding criterion makes additional predictions about the shape and the noise structure of the value function as we change the payoff distribution.

To test our theory, we conduct a laboratory experiment in which subjects make a series of decisions between a risky lottery and a certain option. We experimentally manipulate the distribution of risky payoffs across two conditions: one in which payoffs in the choice set are drawn from a distribution with high volatility, and another in which the distribution has low volatility. We find that, within subjects, risk taking is more sensitive to payoffs in the low

volatility condition, compared to the high volatility condition. This is consistent with our main theoretical prediction, that risk taking is more sensitive to payoffs that are expected to appear in the choice set more frequently.

We also find that after the payoff distribution switches from low volatility to high volatility, the sensitivity to extreme payoffs increases over time. This suggests that the perception of an "outlier payoff" is time-varying: extreme values which begin to appear more frequently are no longer perceived as outliers, and thus are subject to less noisy encoding. Our data indicate that this change in perception occurs relatively quickly, within five trials (Payzan-LeNestour and Woodford, 2018). Conversely, we find evidence that the sensitivity to intermediate payoffs *decreases* over time in the high volatility condition. The interpretation here is that as intermediate values become less frequent (relative to the low volatility condition), perceptual resources are allocated away from these intermediate values. Overall, our experimental results provide novel evidence that the perception of risky payoffs depends on the recently encountered payoff distribution, similar to how our perception of light depends on the environment to which our eyes have recently adapted.

While we formally present the full model later in the paper, we briefly explain the key assumptions and mechanisms here. There are two basic building blocks of our model. First, as in KLW, we assume the decision maker encodes the absolute value of each risky payoff with noise, conditional on the choice set that is perfectly observable to the econometrician. Specifically, when the *DM* is presented with a choice set in which a risky lottery pays $X$ dollars in some state, we assume that the *DM* perceives this payoff as some noisy mental representation, $R_x$, which is governed by a probability density function $p(R_x \mid X)$. This assumption captures a fundamental feature of numerical cognition, which holds that our perception of numerical quantities is noisy, even when these quantities are presented in symbolic form—for example, through Arabic numerals (Dehaene, 2011).

The second building block is that the *DM* understands that numerical quantities are encoded with noise, and unconsciously performs Bayesian inference to form an optimal estimate

of the numerical payoff under consideration. This assumption may appear heroic, but it is guided by the literature on sensory perception which finds a tight link between quantitative predictions from a Bayesian framework and data from controlled experiments (Stocker and Simoncelli, 2006; Girshick, Landy, and Simoncelli, 2011; Dehaene, 2014; Wei and Stocker, 2015; 2017). We assume that the *DM* forms a prior through learning the distribution of previously encountered payoffs, and we endogenize the likelihood function through efficient coding. Interestingly, for some prior distributions, the likelihood functions that we derive exhibit logarithmic compression, which resembles the likelihood functions that are assumed in KLW. A key difference, however, is that our likelihood functions are inextricably linked to the parameters of the payoff distribution to which the *DM* has adapted. Once the *DM* performs Bayesian inference on the payoffs from each lottery, she chooses the lottery with the maximum expected value, conditional on the optimal estimate of each payoff.

The model we present here is meant to capture intuitive judgments about choice under risk, such as the judgments between simple gambles that Kahneman and Tversky (1979) sought to explain with prospect theory. Our model does not apply to all decisions under risk, and in particular, it should not be applied to high stakes decisions that are based on explicit symbolic calculations. These decisions are less affected by the imperfect perception that drives our model, and instead are likely to be governed by a separate decision-making system. At the same time, our model is not necessarily confined to low stakes decisions, and we believe that it is reasonable to apply in situations similar to those where prospect theory has found success (see Barberis 2013 for a review).

Within our model, there are two limitations worth emphasizing. First, the model is not dynamic and thus makes no explicit predictions about how the speed with which the *DM* adapts to new environments affects risk taking. Instead, our model makes precise comparative static predictions, which we use to guide our experimental design. Second, we assume that only payoffs are subject to noisy encoding, but that probabilities are perceived without noise. This assumption is for simplicity, and in reality, state probabilities are also likely to be encoded with noise. To further test whether our experimental results are indeed generated by the noisy

encoding of payoffs, we run an additional experiment in which the subject still needs to perceive numerical quantities, but there is no need to perceive probabilities or integrate them with payoffs. We run a *riskless* choice experiment where we incentivize subjects to classify whether a sequence of numbers is above or below a reference number. We find that even in this simpler environment, accuracy depends on the distribution of numbers to which the subject has adapted. For a given number, subjects exhibit greater classification accuracy if the number has occurred more frequently in the recent past. This provides some support for our basic model assumption that the perception of numerical symbols is noisy and changes across environments.

Our paper contributes to a recent literature that examines the effect of imperfect perception and Bayesian inference on economic choice. Gabaix and Laibson (2017) show theoretically that a *DM* with a discount rate of one will appear impatient if payoffs delivered farther in the future are perceived with more noise. Woodford (2012) and KLW provide a framework in which a *DM* with linear utility can appear risk averse if payoffs are encoded with noise. Steiner and Stewart (2016) show that Bayesian inference can generate an overweighting of small probability events, as in prospect theory.[1] Both of our experiments provide evidence that supports the type of perceptual processes proposed in these Bayesian models of economic choice.

Our results also contribute to a literature that uses attention and basic neural computations to constrain patterns of risky choice (Bordalo, Gennaioli, and Shleifer, 2012; 2017; Tymula and Glimcher, 2017; Landry and Webb, 2018). A particularly relevant neural computation is that of *normalization,* in which the brain normalizes stimulus values according to the distribution of values in the environment. Several experiments have found evidence consistent with normalization in the brain (Tobler, Fiorillo, and Schultz, 2005; Carandini and Heeger, 2012; Rangel and Clithero, 2012; Louie and Glimcher, 2012), but there is less evidence that this process has an associated effect on behavior. Khaw, Glimcher, and Louie (2017) show that the valuation of consumer items negatively correlates with the average value of recently

[1] In related work, Bhui and Gershman (2018) show how efficient coding can provide a normative foundation for a model of multi-attribute decision making called decision by sampling (Stewart, Chater, and Brown, 2006).

encountered items, and Polania, Woodford, and Ruff (2018) show that valuation depends on the entire distribution. Here, we demonstrate that these adaptation effects extend into the domain of risky choice.[2]

The paper proceeds as follows. In Section II, we lay out the basic elements of the model. Section III examines the model's implications. Section IV describes the main experiment of the paper, a risky choice experiment, and discusses its results. Section V follows with a riskless choice experiment. Section VI concludes and suggests directions for future research.

## II. The Model

In this section, we develop a model of risky choice based on efficient coding and Bayesian decoding, following the recent work of KLW and Wei and Stocker (2015, 2017).

### *II.1. Choice environment*

The *DM* faces a choice set that contains two options: a certain option and a risky lottery. The certain option, denoted as ($C$, 1), pays $C$ dollars with certainty. The risky lottery, denoted as ($X$, $p$; 0, 1 – $p$), pays $X$ dollars with a probability $p$ and zero dollars with the remaining probability 1 – $p$. The *DM*'s task is to choose between these two options.

Under expected utility theory, a *DM* with utility $U(\cdot)$ chooses the risky lottery over the certain option if and only if

$$p \cdot U(X) + (1-p) \cdot U(0) \geq U(C). \quad (1)$$

Conditional on $X$, $C$, and $p$, the *DM*'s choice is non-stochastic.

Motivated by the literature on sensory perception, we assume that the *DM* imperfectly perceives the payoffs of $X$ and $C$ (Deheane, 2011; Girshick et al., 2011; Wei and Stocker, 2015).[3]

[2] See Payzan-LeNestour, Balleine, Berrada, and Pearson (2016) for experimental evidence on adaptation to variance, and Zimmermann, Glimcher, and Louie (2018) for evidence on adaptive behavior in monkeys in the realm of risky choice.

[3] Further evidence for this assumption comes from recent experimental work which demonstrates that humans have single neurons that selectively and stochastically respond to "preferred" numbers (Kutter, Bostroem, Elger,

Specifically, before observing the choice set, the *DM* has a prior over the distributions of $X$ and $C$.[4] Upon observing the choice set, the presentations of $X$ and $C$ elicit a noisy sensory representation of $X$, $R_x$, and a noisy sensory representation of $C$, $R_c$, each randomly drawn from a distinct likelihood function; this process of creating a sensory representation of a stimulus value is called *encoding*. The *DM* then uses Bayesian inference to form optimal estimates of $X$ and $C$, $\mathbb{E}[\tilde{X}|R_x]$ and $\mathbb{E}[\tilde{C}|R_c]$; this process of combining priors with the likelihood functions to form posterior beliefs is called *decoding*.[5] As in KLW, we further assume that the *DM* has linear utility, and thus chooses the risky lottery if and only if $p \cdot \mathbb{E}[\tilde{X}|R_x] > \mathbb{E}[\tilde{C}|R_c]$.

It is worth noting that the encoding process described above is conditional on the values of $X$ and $C$, which we assume are perfectly observable to the econometrician. This is different from learning about the realizations of $X$ and $C$ in *future* choice sets. We also note that, because the encoding process is noisy, sensory representations $R_x$ and $R_c$ vary from trial to trial conditional on $X$ and $C$. Given this, the *DM*'s choice between the certain option and the risky lottery is intrinsically stochastic.

*II.2. The likelihood function*

Consider a probability density function of a noisy representation $m$ for a given stimulus value $\theta$. We denote such a function as $p(m|\theta)$. The likelihood function mentioned above is then formally defined as

$$L(\theta \mid m) = p(m \mid \theta). \tag{2}$$

This function governs the likelihood for each stimulus value $\theta$ conditional on the noisy representation $m$. We assume an efficient coding criterion proposed in Wei and Stocker (2015) to constrain the likelihood function. This criterion requires

Mormann, and Nieder, 2018). Such "number neurons" are likely to generate the noisy perception of symbolic numbers.

[4] We assume in this model that the probability $p$ is perceived without noise. In our experimental design in Section IV, we set $p$ to a constant across trials so that, through learning, it is plausible that the *DM* perceives the precise value of $p$.

[5] See von Helmholtz (1925), Curry (1972), and Knill and Richards (1996) for earlier work on sensory perception as a form of Bayesian inference.

$$\sqrt{J(\theta)} \propto p(\theta), \tag{3}$$

where Fisher information $J(\theta)$ is given by

$$J(\theta) = \int \left( \frac{\partial \ell \mathrm{n}\ p(m \mid \theta)}{\partial \theta} \right)^2 p(m \mid \theta) dm, \tag{4}$$

and $p(\theta)$ is the true probability density function of the stimulus value $\theta$.

Intuitively, Fisher information $J(\theta)$ measures the amount of coding resources allocated towards accurate perception of a given stimulus value $\theta$. As a result, the efficient coding condition (3) implies that encoding accuracy is greater for stimulus values that occur more frequently.

To find a likelihood function for a fixed distribution $p(\theta)$ that satisfies the efficient coding condition in (3), we transform the stimulus space into a "sensory space" through a change of variable $\tilde{\theta} = F(\theta)$, where $F(\theta) = \int_{-\infty}^{\theta} p(y) dy$ is the cumulative density function of $\theta$. It is easy to show that the efficient coding condition is satisfied in the sensory space if the transformed likelihood function $L(\tilde{\theta} \mid m)$ is *location-independent*:

$$L(\tilde{\theta} \mid m) = g(\tilde{\theta} - m), \tag{5}$$

where $g(\cdot)$ is some smooth density function that integrates to one.

In this paper, we further assume that $g(\cdot)$ takes the form of a normal probability density function

$$g(\tilde{\theta} - m) \equiv \frac{1}{\sigma} \phi\left( \frac{\tilde{\theta} - m}{\sigma} \right) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left( -\frac{(\tilde{\theta} - m)^2}{2\sigma^2} \right), \tag{6}$$

where $\phi(\cdot)$ stands for the probability density function of a standard normal distribution, $m$ takes the range of $(-\infty, \infty)$, and parameter $\sigma$ represents the amount of coding resources available to the *DM*: a lower $\sigma$ means a larger amount of coding resource available, and therefore a narrower likelihood function.[6]

[6] Our model's implications are qualitatively robust to the assumption that $g(\cdot)$ is normal; other forms of smooth function lead to similar implications.

Given that the *DM*'s objective is to decode in the stimulus space, we need to transform $L(\tilde{\theta} \mid m)$ back to $L(\theta \mid m)$, which simply requires

$$L(\theta \mid m) = g(F(\theta) - m) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left( -\frac{(F(\theta) - m)^2}{2\sigma^2} \right). \tag{7}$$

In the context of our choice environment, we assume that the probability density functions of $X$ and $C$ are lognormal

$$\begin{aligned} p(X; \mu_x, \sigma_x) &= \frac{1}{\sqrt{2\pi} \cdot \sigma_x X} \exp\left( -\frac{(\ell \mathrm{n}\, X - \mu_x)^2}{2\sigma_x^2} \right), \\ p(C; \mu_c, \sigma_c) &= \frac{1}{\sqrt{2\pi} \cdot \sigma_c C} \exp\left( -\frac{(\ell \mathrm{n}\, C - \mu_c)^2}{2\sigma_c^2} \right). \end{aligned} \tag{8}$$

We now construct the likelihood functions for $X$ and $C$. First, the cumulative density functions that we use for the transformation are

$$\begin{aligned} F(X; \mu_x, \sigma_x) &\equiv \int_0^X p(y; \mu_x, \sigma_x) dy = \Phi\left( \frac{\ell \mathrm{n}\, X - \mu_x}{\sigma_x} \right), \\ F(C; \mu_c, \sigma_c) &\equiv \int_0^C p(y; \mu_c, \sigma_c) dy = \Phi\left( \frac{\ell \mathrm{n}\, C - \mu_c}{\sigma_c} \right), \end{aligned} \tag{9}$$

where $\Phi(\cdot)$ stands for the cumulative density function of a standard normal distribution. We then follow (7) to construct the likelihood functions as

$$\begin{aligned} L(X \mid R_x) &= p(R_x \mid X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left( -\frac{(\Phi((\ell \mathrm{n}\, X - \mu_x)/\sigma_x) - R_x)^2}{2\sigma^2} \right), \\ L(C \mid R_c) &= p(R_c \mid C) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left( -\frac{(\Phi((\ell \mathrm{n}\, C - \mu_c)/\sigma_c) - R_c)^2}{2\sigma^2} \right). \end{aligned} \tag{10}$$

The expressions in (10) characterize the likelihood functions for each value of $X$ and $C$ in the stimulus space, and we emphasize that they depend directly on parameters of the stimulus distributions, $\mu_x$, $\sigma_x$, $\mu_c$, and $\sigma_c$.

To illustrate these likelihood functions, we plot in Figure 1 the stimulus distribution of $X$ and the resulting likelihood function $L(X \mid R_x)$ given several values of $R_x$, for a low volatility

distribution ($\sigma_x = 0.15$) and a high volatility distribution ($\sigma_x = 0.4$). To get a sense of the relative magnitude for the values of $R_x$, we also plot the distribution of $R_x$ defined as

$$p(R_x) = \int_0^{\infty} p(R_x \mid X)\, p(X) dX. \tag{11}$$

It is easy to check that, under assumptions of (8) and (10), the shape of $p(R_x)$ does not depend on $\mu_x$ and $\sigma_x$. More broadly, the shape of $p(R_x)$ does not depend on the shape of the stimulus distribution: all continuous stimulus distributions lead to the same $p(R_x)$. Furthermore, as $\sigma$ goes to zero—that is, as the amount of coding resource available goes to infinity—$p(R_x)$ converges to a uniform distribution between zero and one. A more detailed discussion of properties of $p(R_x)$ is left to the Appendix.

**[Place Figure 1 about here]**

Figure 1 highlights some important features of the likelihood function $L(X \mid R_x)$. For a given stimulus distribution—that is, holding $\mu_x$ and $\sigma_x$ fixed—the shape of the likelihood function depends heavily on the sensory representation $R_x$. Moreover, for a fixed value of $R_x$, shifting the stimulus distribution—in particular, shifting $\sigma_x$—alters the shape of the likelihood function. For example, a higher standard deviation of the stimulus distribution results in higher dispersion of the likelihood function.

It is useful to compare the likelihood function we derive to the one assumed in the model of KLW. As discussed above, for a continuous stimulus distribution $p(X)$ over the range of [0, ∞), the cumulative density function is $F(X) = \int_0^X p(y)dy$. The corresponding likelihood function is

$$L(X \mid R_x) = g(F(X) - R_x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left( -\frac{(F(X) - R_x)^2}{2\sigma^2} \right). \tag{12}$$

In comparison, the KLW likelihood function is

$$L_{KLW}(X \mid R_x) = \frac{1}{\sqrt{2\pi}\cdot\nu} \exp\left( -\frac{(\ell \mathrm{n} X - R_x)^2}{2\nu^2} \right), \tag{13}$$

where ν is a positive coefficient. This likelihood function captures the well-established finding of "scalar variability," whereby larger numbers are encoded with more noise than smaller numbers (Dehaene, 2011). When comparing (12) with (13), one can see that the two likelihood functions are, in general, quite different. However, when $F(X)$ and $\ell n X$ have similar shapes, the likelihood function derived from efficient coding will resemble the KLW likelihood function. In particular, because the derivative of $\ell n X$ is $1/X$, which is a monotonically decreasing function, we conjecture that a monotonically decreasing stimulus distribution $p(X)$ will generate a likelihood function that resembles the one assumed in KLW.

In Figure 2, we present a specific example with the stimulus distribution taking a form of a gamma distribution

$$p(X;k,\theta) = \frac{1}{\Gamma(k)\theta^k} X^{k-1} e^{-X/\theta}, \tag{14}$$

where $\Gamma(\cdot)$ is the gamma function, $\theta > 0$ and $k > 0$. We choose parameter values for $\theta$ and $k$ so that $p(X; k, \theta)$ is a monotonically decreasing function.

**[Place Figure 2 about here]**

Figure 2 shows that, with a monotonic stimulus distribution, the likelihood functions in (12) (middle panel) and the KLW likelihood functions (bottom panel) look reasonably similar.[7] One difference, however, is that for right-tail values of $R_x$, efficient coding tends to generate likelihood functions that are more positively skewed compared to the likelihood functions in KLW. More generally, the likelihood functions in KLW are invariant to the parameters $\theta$ and $k$, whereas efficient coding implies that the likelihood functions will depend closely on the entire shape of the stimulus distribution. When comparing Figures 1 and 2, one can see that efficient coding implies very different likelihood functions when the stimulus distribution is lognormal,

[7] Dehaene and Mehler (1992) provide evidence that a monotonically decreasing distribution provides a good approximation of the distribution of numbers in natural settings. This suggests that scalar variability may arise as a consequence of efficient coding when combined with the distribution of naturally occurring numbers. The monotonic decline in the frequency of numbers holds across different cultures, with the exception that there are small increases at round numbers such as 10, 20, and 50.

compared to when it is monotonic. For the rest of the paper, we retain the assumption of (8) that the stimulus distribution is lognormal.

We conclude our discussion on the likelihood function by making a remark on adaptation. Notice first that our model is static. As a result, it does not make explicit predictions about how the likelihood function evolves over time. However, the comparative static predictions of the model shed light on the perception of outliers. Figure 1 above shows that, when the stimulus distribution has a low volatility of $\sigma_x = 0.15$, the value $X = 30$ is perceived as an outlier for frequently occurring values of $R_x$, because it is located in the right tail of the associated likelihood functions. However, when the stimulus distribution has a higher volatility of $\sigma_x = 0.4$, the value $X = 30$ is much less of an outlier as the likelihood functions become more dispersed and they assign higher likelihood to the value $X$ =30. This comparative static result suggests that as the *DM* adapts to a different payoff distribution, her perception of outliers changes.

*II.3. Bayesian decoding*

The likelihood functions derived in the previous section, in conjunction with prior beliefs of the true stimulus distribution, can be used for Bayesian decoding of the realized sensory representations. That is, estimates of $X$ and $C$ conditional on $R_x$ and $R_c$, $\mathbb{E}[\tilde{X}|R_x]$ and $\mathbb{E}[\tilde{C}|R_c]$, are given by

$$\mathbb{E}[\tilde{X}|R_x] = \frac{\int_0^\infty p(R_x \mid X)\, p_0(X)\, X dX}{\int_0^\infty p(R_x \mid X)\, p_0(X) dX} \tag{15}$$

and

$$\mathbb{E}[\tilde{C}|R_c] = \frac{\int_0^\infty p(R_c \mid C)\, p_0(C)\, C dC}{\int_0^\infty p(R_c \mid C)\, p_0(C) dC}, \tag{16}$$

where $p_0(X)$ and $p_0(C)$ are the *DM*'s prior beliefs about $X$ and $C$. Here, we equate these priors with the true stimulus distributions described in (8)

$$p_0(X) = p(X;\mu_x,\sigma_x), \qquad p_0(C) = p(C;\mu_c,\sigma_c). \tag{17}$$

We assume that learning leads the *DM* to adapt her priors to the true stimulus distribution. While we do not explicitly incorporate the adaptation process in our model, we provide some experimental evidence of the speed of adaptation in Section IV.

Because the *DM* has linear utility by assumption, she chooses the risky lottery over the certain option if $p \cdot \mathbb{E}[\tilde{X}|R_x] > \mathbb{E}[\tilde{C}|R_c]$. Conversely, she chooses the certain option if $p \cdot \mathbb{E}[\tilde{X}|R_x] \leq \mathbb{E}[\tilde{C}|R_c]$.

A key feature of our model is that we jointly constrain both the prior and the likelihood functions by the true stimulus distribution. It follows that shifting the true stimulus distribution affects both the encoding and the decoding process. In the next section, we develop the model's implications for risky choice behavior.

## III. Model Implications

In this section, we examine the implications of the model. We begin by studying how the probability of choosing the risky lottery changes with the stimulus payoff distribution. We then look at the model's implication for risk preferences. Finally, we discuss the subjective value function derived from the model.

### *III.1. Probability of risk taking*

Conditional on *X* and *C*, the sensory representations $R_x$ and $R_c$ are drawn from the probability density functions $p(R_x \mid X)$ and $p(R_c \mid C)$. For given $R_x$ and $R_c$, the *DM* chooses between the risky lottery and the certain option based on (15), (16), and (17). Holding *X*, *C*, and the stimulus distribution fixed, we can compute the probability of risk taking—that is, the probability of choosing the risky lottery over the certain option—over many realizations of $R_x$ and $R_c$

$$\mathbb{P}\text{rob}(\textit{risk taking} \mid X, C) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{\{p\cdot\mathbb{E}[\tilde{X}|R_x]>\mathbb{E}[\tilde{C}|R_c]\}}\, p(R_x \mid X)\, p(R_c \mid C)\, dR_x dR_c. \tag{18}$$

To understand the determinants of the probability of risk taking, Figure 3 plots this probability against the natural logarithm of $X$ over $C$, $\ell n(X/C)$, for different volatility levels of the stimulus distribution: $\sigma_x = \sigma_c = 0.4$, 0.8, and 1.5. Specifically, for each volatility level, we set $C$ to $\exp(\mu_c + \frac{1}{2}\sigma_c^2)$ while changing the value of $X$.

[**Place Figure 3 about here**]

Naturally, a higher ratio of $X$ over $C$ increases the attractiveness of the risky lottery and hence increases the probability of risk taking. Notice that, under expected utility theory and no background wealth, the probability of risk taking should be a step function of $\ell n(X/C)$ with a single step at $\ell n(X/C) = \ell n[U^{-1}((U(C) - (1 - p)U(0))/p)/C]$. However, with the Bayesian inference process described above, the probability of risk taking has an *S*-shaped relationship with $\ell n(X/C)$. Moreover, the overall slope of this function is *negatively* related to volatility of the stimulus distribution. That is, risk taking is more sensitive to payoff values in the low volatility condition, compared to the high volatility condition. The intuition is that, lower stimulus volatility increases the encoding accuracy of stimulus values that occur more frequently in this environment relative to the high volatility condition, and hence improves the discriminability of these values.

More generally, (18) shows that the probability of risk taking is a two-dimensional function of $X$ and $C$. Figure 4 plots the probability for two different volatility levels of the stimulus distribution: $\sigma_x = \sigma_c = 0.19$ (low volatility) and $\sigma_x = \sigma_c = 0.55$ (high volatility).

[**Place Figure 4 about here**]

Figure 4 makes it obvious that $\ell n(X/C)$ is not a sufficient statistic of the probability of risk taking. Instead, $X$ and $C$ jointly affect this probability. For instance, with $\sigma_x = \sigma_c = 0.55$, $\mu_x = 3.05$, and $\mu_c = 2.35$, setting $X$ to $\exp(\mu_x + \frac{1}{2}\sigma_x^2) = 24.6$ and setting $C$ to $\exp(\mu_c + \frac{1}{2}\sigma_c^2) = 12.2$ gives $X/C = 2.01$ and a risk-taking probability of 77.7%. On the other hand, setting $X$ to $\exp(\mu_x + 2 \cdot \sigma_x^2) = 38.7$ and setting $C$ to $\exp(\mu_c + 2 \cdot \sigma_c^2) = 19.2$ gives the same ratio of $X/C = 2.01$ but a lower risk-taking probability of 74.3%.

*II.2 Risk preferences*

Here we examine the set of values of $X$ and $C$ for which the *DM* is indifferent between the two options—that is, when our model predicts a 50% probability of risk taking. Because risk taking is a two-dimensional function of $X$ and $C$, we can trace out an "indifference curve." That is, for a given value of $X$, we compute the value of $C$ such that $\mathbb{P}\text{rob}(\textit{risk taking} \mid X, C) = 0.5$.

[**Place Figure 5 about here**]

We plot this indifference curve in Figure 5. With the parameter values of $\mu_x = 3.05$, $\mu_c = 2.35$, $p = 0.59$, $\sigma_x = 0.55$, $\sigma_c = 0.55$, and $\sigma = 0.1$, we find that for $X > 20.7$, $C(X) < X \cdot p$, implying that the *DM* is risk averse. For $X < 13.7$, $C(X) > X \cdot p$, implying that the *DM* is risk seeking. Finally, for $X$ between 13.7 and 20.7, $C(X) \approx X \cdot p$, implying that the *DM* is about risk neutral. For a very large value of $X$, efficient coding implies that it is hard to discriminate between nearby values—which is similar to the logarithmic compression assumed in KLW—and this leads to risk aversion. Conversely, for a very small value of $X$, efficient coding again implies a lack of discriminability between nearby values. However, this now gives rise to risk-seeking behavior.

*II.3 Value function*

The probability density function $p(R_x|X)$ and the posterior belief $\mathbb{E}[\tilde{X} \mid R_x]$ together generate a subjective valuation of $X$

$$v(X) = \int_{-\infty}^{\infty} \mathbb{E}[\tilde{X} \mid R_x] p(R_x \mid X) dR_x. \tag{19}$$

We call $v(X)$ the *DM*'s subjective value function. Given the randomness in the mental representation of a stimulus value, we can also compute the standard deviation of $v(X)$ as

$$\sigma(X) = \left[ \int_0^{\infty} (\mathbb{E}[\tilde{X} \mid R_x])^2 \, p(R_x \mid X) dR_x - v^2(X) \right]^{1/2}. \tag{20}$$

Figure 6 plots, for both $X$ and $C$ and for both $\sigma_x = \sigma_c = 0.19$ (low volatility) and $\sigma_x = \sigma_c = 0.55$ (high volatility), the subjective values $v(X)$ and $v(C)$, as well as their one-standard-deviation bounds $v(X) \pm \sigma(X)$ and $v(C) \pm \sigma(C)$.

[**Place Figure 6 about here**]

Figure 6 leads to several observations. First, consistent with prospect theory (Kahneman and Tversky, 1979), the lack of discriminability among outliers generates diminishing sensitivity: the marginal utility $v'(X)$ decreases as $X$ becomes either extremely large or extremely small. Second, diminishing sensitivity is more pronounced when stimulus volatility is lower. In this case, a wider range of stimulus values become outliers, and are therefore difficult to discriminate. Third, both the shape of the utility function and the randomness in utility come from noisy encoding: for very large values of $X$, low discriminability leads to both lower marginal utility $v'(X)$ and higher randomness in utility $\sigma(X)$. Finally, the value of $X$ that has the highest slope for $v(X)$, which typically corresponds to the "reference point" in prospect theory, arises endogenously in our framework. Here, it corresponds to the stimulus value that has the highest degree of local discriminability.

## IV. An Experimental Test

In this section we provide an experimental test of our model. Our experiment is designed specifically to test whether risk taking varies with the payoff distribution that the subject encounters.

### *IV.1. Design*

On each trial in the experiment, subjects choose between a risky lottery and a certain option. The risky lottery delivers a positive payoff $X$ with probability $p$, and zero otherwise. The certain option delivers a positive payoff $C$ with certainty. The experiment consists of eight blocks, with sixty trials in each block. Each subject therefore completes a total of four hundred eighty trials, which we index by $t = 1, 2, \ldots, 480$. At the end of the experiment, subjects are paid according to their decision on one randomly selected trial.

We experimentally manipulated the distribution from which payoffs in the choice set are drawn. On each trial, the values of $X$ and $C$ were jointly drawn from a lognormal distribution

$$\begin{pmatrix} \ell \mathrm{n}\, X \\ \ell \mathrm{n}\, C \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mu_x \\ \mu_c \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix} \right). \tag{21}$$

We set the mean values to $\mu_x = 3.05$ and $\mu_c = 2.35$, so that on average, the risky lottery offers a higher expected value than the certain option. Our treatment variable is the standard deviation, which we varied across two conditions: high volatility and low volatility. In the high volatility condition, we set $\sigma = 0.55$, and in the low volatility condition, we set $\sigma = 0.19$. The first block of the experiment was a high volatility block, and the blocks alternated deterministically, so that the experimented ended with a low volatility block (Figure 7). We set the correlation between $\ell$n($X$) and $\ell$n($C$) at $\rho = 0.5$. Although this positive correlation is not part of the model we developed in the previous section, it helped to reduce the number of trivial choice sets where $X < C$ (and as a result, the certain option stochastically dominates the risky lottery). The values of $X$ and $C$ were drawn from their associated distribution (high volatility or low volatility) at the subject-trial level, and thus each subject faced a unique path of payoffs during the experiment.

[**Place Figure 7 about here**]

For all trials, we set the probability that the risky lottery paid $X$ to $p = 0.59$. Following KLW, we chose this design feature for two reasons. First, we used a "non-round" number so that subjects could not easily compute the expected value of the risky lottery—which was more likely to happen if we used, for example, $p = 0.5$ or $p = 0.6$. Second, even though our model assumes that the subject does not encode the probability $p$ with noise, in reality, this variable is also likely to be encoded with noise. By presenting the same value of 0.59 on each trial, this increased the plausibility of our simplifying assumption that subjects precisely encoded this variable. Later in the paper, we conduct an additional experiment to directly test the noisy encoding of payoffs, without appealing to any assumptions about probability encoding.

Before the experiment began, subjects were told that they would be asked to choose between two lotteries on each of four hundred eighty trials and these trials would be separated into eight parts. However, subjects were not given any information about the distribution of $X$ and $C$, nor were they told that these distributions changed across blocks. The exact instructions that were given to subjects before the experiment are provided in the Appendix.

*IV.2. Experimental procedures*

We recruited $N = 34$ subjects for this experiment, which was conducted across three sessions at Caltech and USC. Before starting the experiment, subjects went through a set of ten practice trials to become familiar with the task and the software. Figure 7 shows an example trial from the experiment, in which the risky lottery is presented on the left as a colored bar chart, and the value $X$ is displayed at the bottom next to its associated probability of 0.59. The certain option is presented on the right side of the screen. On each trial, subjects were instructed to select the left or right option by pressing one of two keys. The location of the risky lottery was randomized across subjects and trials, and subjects had unlimited time to make their decision on each trial. At the end of each block of sixty trials, a progress screen appeared, which reported how many of the eight blocks the subject had completed.

At the end of the eighth block, the computer randomly selected one of the four hundred eighty trials from the experiment. If the subject chose the risky lottery on this trial, a random number generator determined whether the subject received the payoff of $\$X$ or the payoff of \$0, according to the probabilities associated with these payoffs. If the subject chose the certain option, she received the amount of $\$C$. In addition to the earnings from this randomly selected trial, each subject received a \$7 show-up fee. The average earning, including the show-up fee, was \$25.89.

*IV.3. Experimental results*

*IV.3.A. Treatment effects*

Subjects chose the risky lottery on 40.5% of trials in the low volatility condition and on 42.7% of trials in the high volatility condition. One subject did not exhibit any variation in risk taking in the low volatility condition (choosing the certain option on each trial), and we exclude this subject from all subsequent analyses.

[**Place Figure 8 about here**]

Figure 8 plots the proportion of trials on which subjects chose the risky lottery, as a function of the natural logarithm of $X$ over $C$, $\ell n(X/C)$. Recall that the probability $p$ stays constant across all trials, and thus $\ell n(X/C)$ provides a good—though insufficient—statistic that summarizes the attractiveness of the risky lottery relative to the certain option. The figure shows that risk taking increases in $\ell n(X/C)$ in both conditions, which provides a basic consistency check on the data. One can also see that the slope of the curve in the low volatility condition appears to be steeper than that in the high volatility condition. This is consistent with a basic prediction of our model: when the stimulus distribution becomes more concentrated, choice sensitivity increases.

To conduct formal empirical tests, we run regressions where the dependent variable takes the value of one (zero) if the subject chose the risky lottery (certain option) on trial *t.* We pool all 15,840 trials across subjects and conditions, and run a logistic regression. The results in Column (1) of Table 1A show that the regression coefficient on $\ell n(X/C)$, which provides a measure of the sensitivity of risk taking in the low volatility condition, is positive and strongly significant. *high* is a dummy variable that takes the value of one if the trial is in the high volatility condition, and zero otherwise. The coefficient of interest is on the interaction term $\ell n(X/C)\times high$, which is significantly negative, indicating that risk taking becomes less sensitive to $\ell n(X/C)$ in the high volatility condition. This provides formal support for a difference in choice sensitivity between the high and low volatility conditions.

[**Place Table 1 about here**]

Our model predicts that this difference in choice sensitivity stems from different *perceptions* of $X$ and $C$ across conditions. However, the distributions of $X$ and $C$ themselves vary across conditions, and thus the difference we detect may simply be driven by a different response to extreme values of $\ell n(X/C)$. In particular, the range of $\ell n(X/C)$ in the low volatility condition is (0.02, 1.36), but in the high volatility condition, there are many trials for which $\ell n(X/C)$ falls outside this range.

To help address this concern, we restrict our regression to trials with similar levels of $\ell n(X/C)$ across the two volatility conditions. We re-estimate the regression in Column (1) using only trials for which $0.02 < \ell n(X/C) < 1.36$; this represents the set of values of $\ell n(X/C)$ which appear in both conditions of our experiment. Column (2) shows that our results are quite similar on this subset of data. Given this, it is unlikely that differences in the current choice set drive the full effect; however, we note that this is not a perfect control because the distribution of $\ell n(X/C)$ still differs within this restricted domain. Columns (3) and (4) show that our main result holds within each half of a block. Column (3) is estimated using data from the first half of each block (the first thirty trials), while column (4) uses data from the second half (the last thirty trials), and the main result holds in both of these subsets of the data.

Column (5) shows our results hold even within the first ten trials of each block. In fact, if we restrict to the *first trial* of each block (Column (6)), we find that the coefficient on the interaction term remains significantly negative (*p*-value of 0.046). This result is potentially concerning, because it is consistent with a theory where the subject adapts to the new distribution instantly. However, another more plausible explanation is based on the perception of outliers. On the first trial of each block of the high volatility condition, the subject has just faced sixty choice sets from the low volatility condition, and presumably has adapted to a narrow range of $\ell n(X/C)$. On the first trial of the high volatility block, there is a good chance that $\ell n(X/C)$ falls outside this range (or near the extremes of this range), and thus this value will be perceived as an outlier.[8] Because the subject's coding resources do not adjust immediately following the low volatility block, the perception of the outlier will be noisy.[9]

[8] In our experiment, there is a 21% chance that $\ell n(X/C)$ falls outside the range (0.02, 1.36) on the first trial of a high volatility block.

[9] We note that in each specification in Table 1A, the coefficient on the *high* dummy variable is significantly positive. This is likely because $\ell n(X/C)$ is not a sufficient statistic for risk taking and may contribute to model misspecification. Table 1B presents regression results where we separately enter the *X* and *C* regressors, as well as their associated interaction terms. We see that the coefficient on the *high* dummy variable is no longer significant in these regressions; moreover, consistent with the result in Table 1A, we also find that risk taking becomes less sensitive to both *X* and *C* in the high volatility condition.

*IV.3.B. Adaptation and dynamics*

The above logic can be extended more generally to all trials in the high volatility block. Consider a value of $\ell n(X/C)$ that is extreme within the context of the low volatility block, but not within the high volatility block. If the subject encounters this value *early* in a high volatility block, it will be perceived as an outlier, but the presentation of this value should lead to a subsequent adjustment of coding resources through adaptation. When a value of $\ell n(X/C)$ that was perceived as extreme early in a high volatility block is presented a few times, then later in the high volatility block, it will no longer be perceived as an outlier because coding resources have been allocated to a wider range of payoffs. Therefore, risk taking should become more sensitive to extreme values over the course of the high volatility block.

To test this prediction, we examine how risk taking varies in the presence of outliers, over the course of the high volatility block. We define an outlier as a value of $\ell n(X/C)$ that is more than three standard deviations from the mean. The standard deviation and mean are computed at the subject-block level, using the sixty trials from the immediately preceding low volatility block (we restrict analysis to blocks 3, 5 and 7, so that each of these high volatility block can be matched to an immediately preceding low volatility block). Under this definition, 30.4% of trials in the high volatility block are considered outliers.

We define a dummy called *outlier* that takes the value of one if the value of $\ell n(X/C)$ is an outlier, and zero otherwise. We also define a dummy called *second* which takes the value of one if the trial takes place in the second half of the block (i.e., in the last thirty trials). We then estimate the following logistic regression, using only data from the high volatility blocks:

$$\begin{aligned} \text{risky}_t = \alpha + \beta_1 \cdot \ell n\left(\frac{X_t}{C_t}\right) + \beta_2 \cdot \ell n\left(\frac{X_t}{C_t}\right) \times outlier \\ + \beta_3 \cdot \ell n\left(\frac{X_t}{C_t}\right) \times second + \beta_4 \cdot \ell n\left(\frac{X_t}{C_t}\right) \times second \times outlier + \varepsilon_t . \end{aligned} \tag{22}$$

In Column (1) of Table 2, the regression coefficient $\beta_1$ on $\ell n(X/C)$ provides the sensitivity of risk taking among intermediate value (non-outlier) trials in the first half of the high volatility block.

As expected, this coefficient is significantly positive. Furthermore, the coefficient $\beta_2$ on the interaction $\ell n(X/C)\times$*outlier* is significantly negative. This indicates that in the first half of the block, shortly after experiencing sixty low volatility trials, subjects are less sensitive to outliers than to intermediate values.

[**Place Table 2 about here**]

The next two rows allow these coefficients to differ in the second half of the block. The coefficient $\beta_4$ on $\ell n(X/C)\times$*second*$\times$*outlier* provides the change in sensitivity of risk taking among outlier trials from the first to second half of the block. Here, we expect the coefficient, $\beta_4$, to be positive, because exposure to outliers during the first half of the high volatility block should trigger adjustment of coding resources to these extreme values in the second half of the block. Conversely, the coefficient $\beta_3$ on $\ell n(X/C)\times$*second* provides the change in sensitivity of risk taking on intermediate value trials between the first and second halves of the block. As the *DM* allocates more coding resources to outliers over the course of the high volatility block, the corresponding resources allocated to intermediate values decrease; we thus expect this coefficient, $\beta_3$, on $\ell n(X/C)\times$*second* to be negative.

We find that the signs of these two coefficients, $\beta_3$ and $\beta_4$, are consistent with these predictions, but neither coefficient is significantly different from zero. To allow for the possibility that adaptation take place more quickly than the first thirty trials, Columns (2) – (4) use more restricted subsets of the data. After restricting to the first and last ten trials of each block in Column (3), we find significant results on both coefficients. We find similar results in Column (4), which suggests that adaptation takes place on the order of five trials, consistent with recent work on adaptation speeds in perceptual decision-making (Payzan-LeNestour and Woodford, 2018).

*IV.3.C. Heterogeneity across subjects*

[**Place Figure 9 about here**]

The data presented in Figure 8 are pooled across subjects, and therefore mask any potential heterogeneity in the change in risk taking across volatility conditions. To investigate this potential heterogeneity, for each subject and condition, we run the following logistic regression:

$$\text{risky}_t = \alpha + \beta \cdot \ell\text{n}\left(\frac{X_t}{C_t}\right) + \varepsilon_t. \tag{23}$$

We record the estimates $\hat{\beta}$ for each subject in the high and low conditions, and plot these against each other in Figure 9. We see there is substantial heterogeneity across subjects in the sensitivity to $\ell$n($X$/$C$). Moreover, subjects who are more sensitive in the high condition are also more sensitive in the low condition. Most importantly, we see that, for a majority of subjects, the data lie above the blue forty-five degree line. This confirms that the increase in choice sensitivity in the low volatility condition is present within most of our subjects.

*IV.3.D. Assumptions about noiseless encoding process*

All of the results in our theoretical model are driven by the noisy encoding of $X$ and $C$. In particular, we make two simplifying assumptions: *i*) there is no noise in encoding the probability $p$; and *ii*) there is no noise in computing the product of $p$ and $\mathbb{E}[X \mid R_x]$ (which is used as the estimate of the expected value of the risky lottery). In reality, there is likely to be noise in both of these processes, which could potentially be responsible for some of the above experimental results. However, because the noisy encoding of payoffs is sufficient to generate our main theoretical predictions, we should still find evidence that the perception of $X$ depends on the recent stimulus distribution, even when there is no need to perceive $p$. To investigate this, we run a follow-up experiment in which the subject still needs to perceive $X$, but does not need to perceive $p$ or integrate probabilities with payoffs.

## IV. Riskless Choice Experiment

*IV.1. Experimental design*

The design of our second experiment is informed by decades of work from the literature on perception of numerical quantities (Moyer and Landauer, 1967). Our experimental design builds on that of Dehaene, Dupoux, and Mehler (1990), who on each trial of their experiment, present subjects with an Arabic numeral between 31 and 99. The subject's task is simply to classify whether the Arabic numeral presented on the screen is larger or smaller than the reference level of 65. Their main result is that as the stimulus numeral gets closer to the reference level, response times increase and classification accuracy decreases. These results are consistent with the noisy encoding of Arabic numerals, which lies at the foundation of both our model of risk taking and that of KLW.

One notable feature of the Dehaene, Dupoux, and Mehler (1990) experiment is that the stimulus distribution is held constant throughout the experiment. Here, we exogenously vary the stimulus distribution, in much the same way that we varied the distribution of monetary amounts in our previous experiment. We have a high volatility distribution (uniform over integers in the range [31, 99]) and a low volatility distribution (uniform over integers in the range [51, 79]). Subjects are incentivized to correctly classify whether each Arabic numeral, which we denote by $X$, is larger or smaller than 65, over sixteen blocks of trials. The blocks alternate between the high volatility condition and the low volatility condition. Each block consists of eighty trials, for a total of 1,280 trials per subject (Figure 10).

[**Place Figure 10 about here**]

We pay subjects as a function of both their accuracy and their speed. In addition to a $7 participation fee, subjects earn a payoff of $(20×*accuracy* – 10×*avgseconds*), where *accuracy* is the percentage of correctly classified trials, and *avgseconds* is the average response time (in

seconds) across all trials in the experiment.[10] In this design, the subject still needs to perceive the value *X*, but there are no probabilities to encode, nor any need to integrate probabilities with payoffs. Therefore, this design provides a clean setting in which we can test whether the perception of an Arabic numeral, *X*, depends on the recently observed stimulus distribution.

*IV.2. Experimental procedures*

We recruited an additional $N = 13$ subjects from Caltech for this experiment. Before the first block, subjects went through a set of ten practice trials to become familiar with the task. On each trial, the stimulus numeral is displayed in white font against a black background, on the center of the screen (Figure 10). Subjects were instructed to press one of two keys to indicate whether the stimulus is smaller or greater than 65. After responding on each trial, a white fixation cross appeared for 500 milliseconds, followed by the stimulus from the next trial. At the end of each block of eighty trials, a progress screen appeared, which reported how many of the sixteen blocks remained. The progress screen was self-paced, and subjects were given the opportunity to take a break during this screen. The average earning, including the show-up fee, was $20.58.

*IV.3. Experimental results*

Subjects accurately classified the stimuli on 90.4% of trials with an average response time of 0.45 seconds. Figure 11 shows the proportion of trials that subjects classified the stimulus as greater than 65, for each value of *X*. If subjects had accurately classified all stimuli, the figure would generate a step function, with a single step at $X = 65$. Instead, the figure replicates results from several previous experiments in the literature, which show that errors decrease in the distance between the two numbers under comparison (Moyer and Landauer, 1967; Dehaene, Dupoux, and Mehler, 1990). To be clear, while it is unsurprising that subjects make errors in general, the more important result is that the error rate is correlated with the distance between the stimulus number and the reference level of 65. It is also worth noting that

[10] From an experimental economics perspective, the fact that our design is incentive compatible provides a small methodological improvement over similar past experiments in the literature on numerical cognition.

the average subject from Caltech has very high mathematical aptitude, and thus the error rates reported here are likely to be close to a lower bound for the error rates among other samples.[11]

[**Place Figure 11 about here**]

Turning to a comparison of our two experimental conditions, we find that subjects correctly classify stimuli on 91.4% of trials in the high volatility condition, and on 89.4% of trials in the low volatility condition. A more informative statistic is the difference in accuracy between conditions, when restricting to stimuli that are common to both conditions: $51 \leq X \leq 79$. This controls for the fact that, on average, trials in the high volatility condition are "easier," in the sense that the average distance to the reference level is greater than in the low volatility condition. We find that accuracy among these trials in the high volatility condition is 86.5%, which is significantly lower than the 89.4% accuracy in the low volatility condition ($p$-value $=$ 0.004). This is consistent with the efficient coding hypothesis: in the low volatility condition, subjects adapt and devote more coding resources to the concentrated range $51 \leq X \leq 79$. In the high volatility condition, subjects need to "spread" these coding resources over a wider range, which leads to increased noise when encoding stimuli in the concentrated range (relative to the low volatility condition).

A sharper test of the efficient coding hypotheses is to compare the slopes in Figure 10. As in our previous experiment, we expect a steeper slope in the low volatility condition. The figure provides suggestive visual evidence for a difference in slopes, and to formally test this, we run a series of logistic regressions. The dependent variable in our logistic regression takes on the value of one if the subject classified $X$ as above 65, and zero otherwise. Column (1) of Table 3 shows that the coefficient on $\ell n(X/65)$ is significantly positive, which indicates that subjects' propensity to classify $X$ as greater than 65 is increasing in $\ell n(X/65)$. More importantly, we find that the coefficient on the interaction term, $\ell n(X/65) \times high$, is significantly negative, indicating that choices are noisier on trials in the high volatility condition.

[11] There is evidence that accurate perception of *non-symbolic* representations of numbers (e.g., a visual array of dots) is positively correlated with mathematical aptitude (Halberda, Mazzocco, and Feigenson, 2008), though it is unclear whether this correlation extends to tasks like ours that use symbolic numerical representations.

[**Place Table 3 about here**]

To control for the difference in distributions of $X$, we re-estimate the regression using data only in the range $51 \leq X \leq 79$. When restricting to this range, the distribution of $X$ on the current trial is the same across conditions, and the only difference is the distribution of previously encountered stimuli. Column (2) provides these estimation results, and we find the slope remains steeper in the low volatility condition (though the difference in slopes is smaller compared to the estimates using the full sample in Column (1)).

One assumption we make in interpreting the results in these first two columns, is that there is no "external stimulus" noise: the stimulus number is displayed clearly on the screen and the font is easy to read (as opposed to, e.g., fuzzy text). We assume that the noise that corrupts the mental representation of the stimulus is based on the internal noise in the subject's nervous system. Nonetheless, it is plausible that comparing 59 with 65 may be easier than comparing 60 with 65, not because of the distance, but because the first digits are visually distinct. To address this, we re-estimated the regression in Column (1) using only trials for which the first digit differs from the first digit of the reference level: $\{X < 60, X \geq 70\}$. Column (3) shows that the slope in the low volatility condition remains steeper, indicating that such a "first-digit" effect cannot explain the full extent of the shift in slope.

To summarize this experiment, we find that the accuracy of classifying an Arabic numeral is affected by *i*) the distance to a reference level, and *ii*) the distribution of previously encountered stimuli. The latter result provides useful evidence supporting a basic assumption of our model of risky choice. Specifically, in an experimental task where there is no need to encode probabilities or integrate with payoffs, we find that choice sensitivity depends on the distribution of previously encountered stimuli.

## IV. Conclusion

In this paper, we derive the implications for risk taking when the perception of payoffs is noisy and governed by efficient coding. Our main theoretical contribution is to endogenize the

likelihood function in the KLW framework, which can normatively justify why some payoffs are encoded with more noise than others. Earlier work by Woodford (2012) provides a model of efficient coding of risky payoffs, but in that model, imperfect perception is applied to the net gain of a payoff. The model of KLW that we build on here instead assumes that the *DM* encodes the absolute value of a symbolic number. This is a more realistic assumption, as the perceptual system responsible for noisy encoding of numerosity is unlikely to support negative numbers (Feigenson, Dehaene, and Spelke, 2004).[12]

To test our model, we conduct two laboratory experiments in which we find evidence consistent with efficient coding of risky payoffs. Specifically, risk taking becomes more sensitive to those payoffs that appear more frequently in the choice set. Such adaptation takes place relatively quickly, on the order of five experimental trials. In our second experiment, where subjects need only classify whether a stimulus is larger than a reference level, we find that classification accuracy changes with the distribution of recently experienced numbers. This provides evidence supporting our basic model assumption, that perception of numerical payoffs is noisy and changes across environments. In our model, this optimal change in perception is driven by the entire stimulus distribution, but in reality, this process can only be approximately implemented due to computational and data constraints. A better understanding of which features of the payoff distribution are most important in driving adaptation is an important topic for future work.

[12] As in KLW, negative numbers can be accommodated in our model by first encoding the absolute value, and then performing multiplication.

# References


Barberis, Nicholas, "Thirty Years of Prospect Theory in Economics: A Review and Assessment," *Journal of Economic Perspectives*, 27 (2013), 173–196

Barlow, Horace, "Possible Principles Underlying the Transformations of Sensory Messages," in *Sensory Communication* (ed. Rosenblith, W.A.) 217–234, MIT Press, 1961.

Bhui, Rahul and Samuel Gershman, "Decision by Sampling Implements Efficient Coding of Psychoeconomic Functions," *Psychological Review*, (2018), forthcoming.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience Theory of Choice Under Risk," *Quarterly Journal of Economics*, 127 (2012), 1243–1285.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Memory, Attention, and Choice," NBER working paper No. 23256, (2017).

Carandini, Matteo, and David J. Heeger, "Normalization as a Canonical Neural Computation," *Nature Reviews Neuroscience*, 13 (2012), 51–62.

Curry, Renwick E., "A Bayesian Model for Visual Space Perception," In *Seventh Annual Conference on Manual Control. NASA SP-281*, Page 187ff, Washington D.C., 1972. NASA.

Dehaene, Stanislas, *The Number Sense*, Oxford: Oxford University Press, 2011.

Dehaene, Stanislas, *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, New York: Viking Press, 2014.

Dehaene, Stanislas, Emmanuel Dupoux, and Jacuqes Mehler, "Is Numerical Comparison Digital? Analogical and Symbolic Effects in Two-Digit Number Comparison," *Journal of Experimental Psychology: Human Perception and Performance*, 16 (1990), 626–641.

Dehaene, Stanislas and Jacque Mehler, "Cross-Linguistic Regularities in the Frequency of Number Words," *Cognition*, 43 (1992), 1–29.

Feigenson, Lisa, Stanislas Dehaene, and Elizabeth Spelke, "Core Systems of Number," *Trends in Cognitive Science*, 8 (2004), 307–314.

Gabaix, Xavier, and David Laibson, "Myopia and Discounting," NBER working paper No. 23254, (2017).

Girshick, Ahna R., Michael S. Landy, and Eero P. Simoncelli, "Cardinal Rules: Visual Orientation Perception Reflects Knowledge of Environmental Statistics," *Nature Neuroscience*, 14 (2011), 926–932.

Halberda, Justin, Michele Mazzocco, and Lisa Feigenson, "Individual Differences in Non-Verbal Number Acuity Correlate with Maths Achievement," *Nature*, 455 (2008), 665–668.

Imas, Alex, "The Realization Effect: Risk-Taking after Realized versus Paper Losses," *American Economic Review*, 106 (2016), 2086–2109.

Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47 (1979), 263–292.

Khaw, Mel Win, Paul W. Glimcher, and Kenway Louie, "Normalized Value Coding Explains Dynamic Adaptation in the Human Valuation Process," *Proceedings of the National Academy of Sciences* 114, (2017), 12696–12701.

Khaw, Mel Win, Ziang Li, and Michael Woodford, "Cognitive Imprecision and Small-Stakes Risk Aversion," NBER working paper No. 24978, (2018).

Knill, David C., and Whitman Richards, *Perception as Bayesian Inference*, Cambridge University Press, 1996.

Kutter, Esther, Jan Bostroem, Christian Elger, Florian Mormann, and Andreas Nieder, "Single Neurons in the Human Brain Encode Numbers," *Neuron* 100, (2018), 1–9.

Landry, Peter, and Ryan Webb, "Pairwise Normalization: A Neuroeconomic Theory of Multi-Attribute Choice," SSRN working paper, (2018).

Laughlin, Simon, "A Simple Coding Procedure Enhances a Neuron's Information Capacity," *Zeitschrift fur Naturforschung* 36, (1981), 9–10.

Louie, Kenway, and Paul W. Glimcher, "Efficient Coding and the Neural Representation of Value," *Annals of the New York Academy of Sciences* 1251, (2012), 13–32.

Moyer, Robert S., and Thomas K. Landauer, "Time Required for Judgments of Numerical Inequality," *Nature* 215, (1967), 1519–1520.

Payzan-LeNestour, Elise, Bernard Balleine, Tony Berrada, and Joel Pearson, "Variance After-Effects Distort Risk Perception in Humans," *Current Biology* 26, (2016), 1500–1504.

Payzan-LeNestour, Elise, and Michael Woodford, "'Outlier Blindness': Efficient Coding Generates an Inability to Represent Extreme Values," SSRN working paper, (2018).

Polania, Rafael, Michael Woodford, and Christian Ruff, "Efficient Coding of Subjective Value," *bioRxiv*, (2018).

Rangel, Antonio, and John A. Clithero, "Value Normalization in Decision Making: Theory and Evidence," *Current Opinion in Neurobiology* 22, (2012), 970–981.

Steiner Jakub, and Colin Stewart, "Perceiving Prospects Properly," *American Economic Review*

106, (2016), 1601–1631.

Stewart, Neil, Nick Chater, and Gordon Brown, “Decision by Sampling,” *Cognitive Psychology* 53, (2006), 1–26.

Stocker Alan A., and Eero P. Simoncelli, “Noise Characteristics and Prior Expectations in Human Visual Speed Perception,” *Nature Neuroscience* 9, (2006), 578–585.

Thaler, Richard, and Eric Johnson, “Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice,” *Management Science* 36, (1990), 643–660.

Tobler Philippe N., Christopher D. Fiorillo, and Wolfram Schultz, “Adapting Coding of Reward Value by Dopamine Neurons,” *Science* 307, (2005), 1642–1645.

Tymula, Agnieszka, and Paul Glimcher, “Expected Subjective Value Theory (ESVT): A Representation of Decision Under Risk and Certainty,” SSRN working paper, (2017).

von Helmholtz HLF, *Treatise on Physiological Optics*, New York: Dover, 1925.

Weber, Martin, and Colin Camerer, “The Disposition Effect in Securities Trading: An Experimental Analysis,” *Journal of Economic Behavior and Organization* 33, (1998), 167–184.

Wei, Xue-Xin, and Alan A. Stocker, “A Bayesian Observer Model Constrained by Efficient Coding Can Explain ‘Anti-Bayesian’ Percepts,” *Nature Neuroscience* 18, (2015), 1509–1517.

Wei, Xue-Xin, and Alan A. Stocker, “Lawful Relation between Perceptual Bias and Discriminability,” *Proceedings of the National Academy of Sciences* 114, (2017), 10244–10249.

Woodford, Michael, “Prospect Theory as Efficient Perceptual Distortion,” *American Economic Review Papers & Proceedings* 102, (2012), 41–46.

Zimmermann, Jan, Paul Glimcher, and Kenway Louie, “Multiple Timescales of Normalized Value Coding Underlie Adaptive Choice Behavior,” *Nature Communications* 9, (2018), 3206.

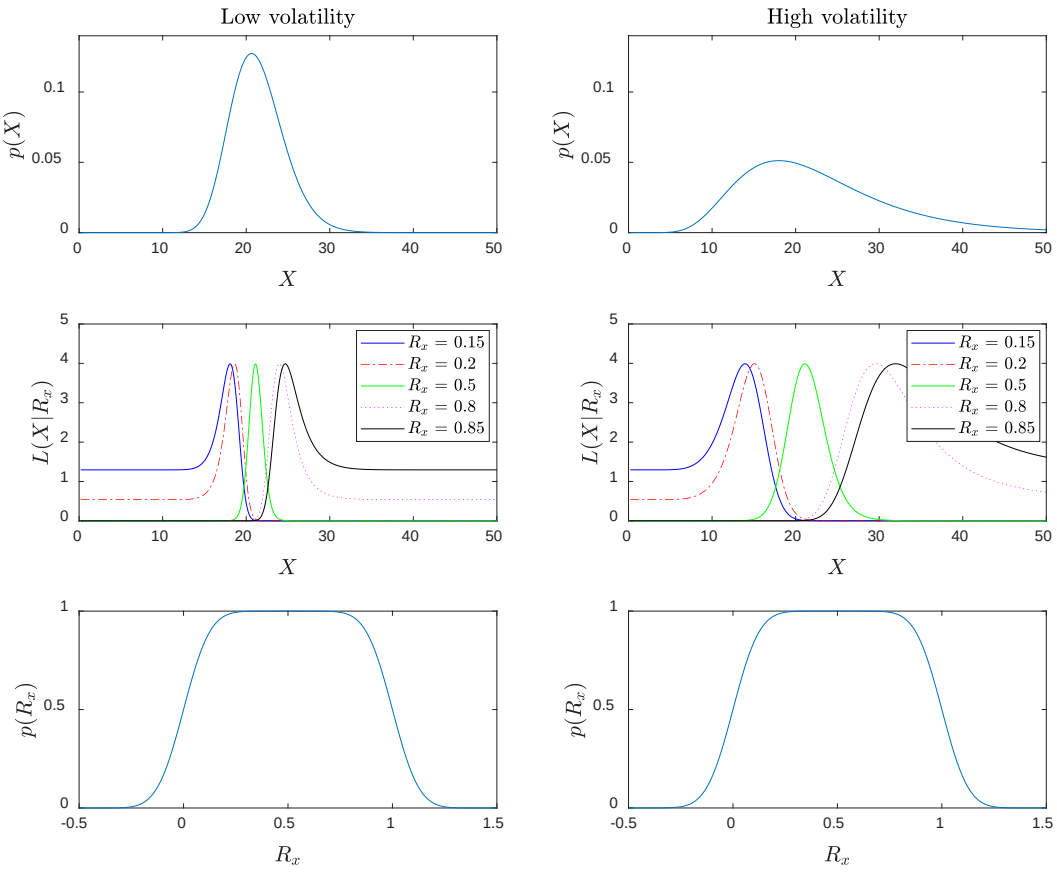


**Figure 1. The stimulus distribution and the likelihood function.** The upper graph plots the stimulus distribution of $X$ according to (8). The middle graph plots the resulting likelihood function $L(X|R_x)$ according to (10), for $R_x = 0.15$, 0.2, 0.5, 0.8, and 0.85. The lower graph plots the density function for $R_x$ according to (11). The parameter values are: $\mu_x = 3.05$ and $\sigma = 0.1$. For the low volatility condition (plots on the left), $\sigma_x = 0.15$. For the high volatility condition (plots on the right), $\sigma_x = 0.4$.

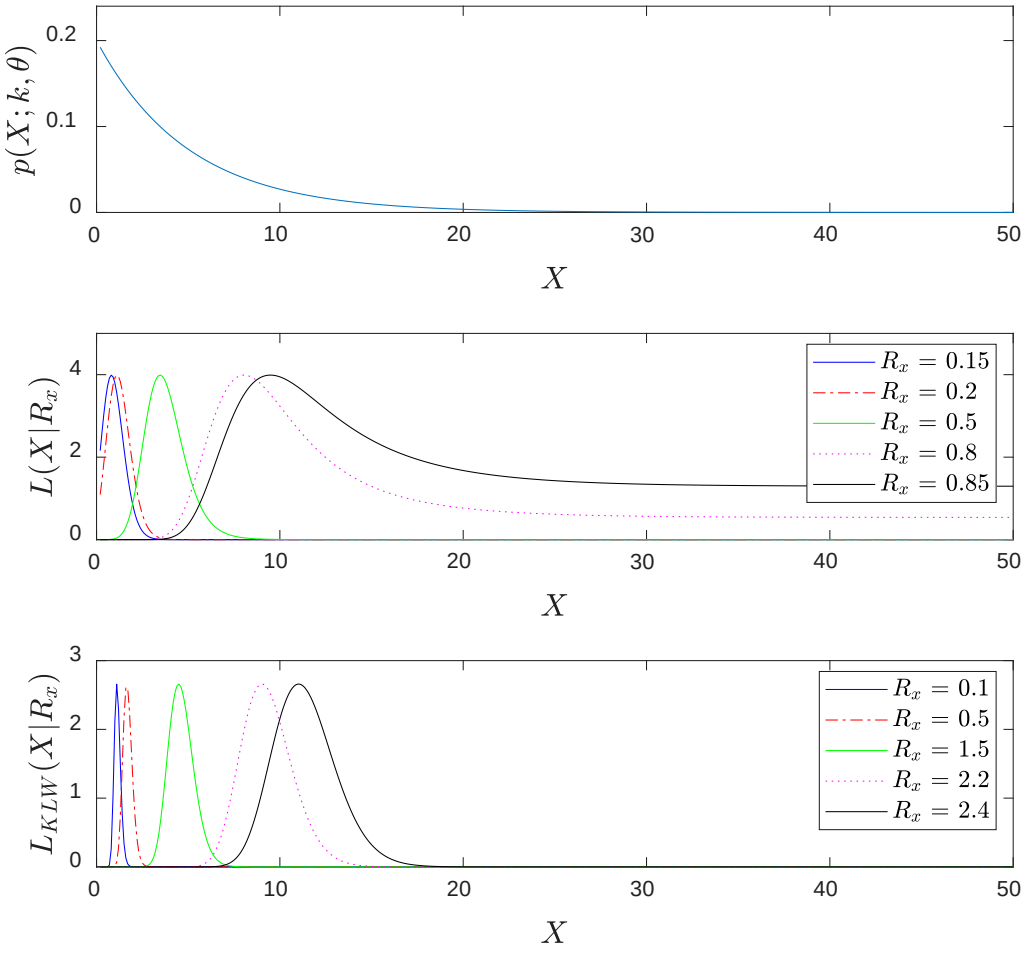


**Figure 2. A monotonic stimulus distribution and the resulting likelihood functions with increasing dispersion.** The upper graph plots a stimulus distribution of $X$ taking the form of a gamma distribution in (14). The middle graph plots the likelihood function implied by efficient coding, according to (12), for $R_x = 0.15$, 0.2, 0.5, 0.8, and 0.85. The lower graph plots the KLW likelihood function in (13), for $R_x = 0.1$, 0.5, 1.5, 2.2, and 2.4. The parameter values are: $k = 1$, $\theta = 5$, $\sigma = 0.1$, and $\nu = 0.15$.

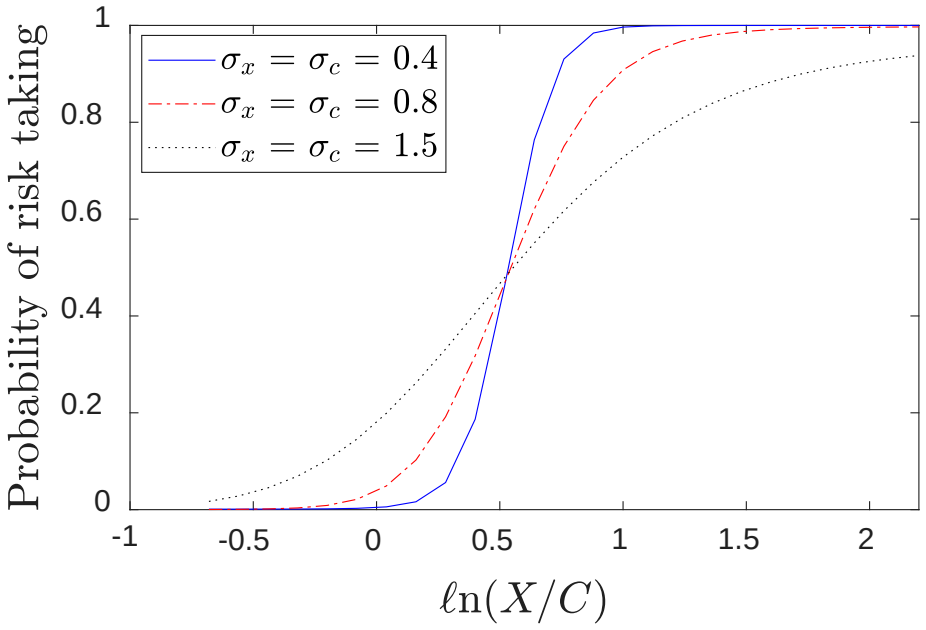


**Figure 3. The probability of risk taking as a function of $\ell$n($X$/$C$).** The figure plots, for each volatility level of the stimulus distribution, $\sigma_x = \sigma_c = 0.4$, 0.8, and 1.5, the probability of risk taking computed in (18) against the natural logarithm of $X$ over $C$, $\ell$n($X$/$C$). Specifically, for each volatility level, $C$ is set to $\exp(\mu_c + \frac{1}{2}\sigma_c^2)$ while we change the value of $X$. The other parameter values are: $\mu_x = 3.05$, $\mu_c = 2.35$, $p = 0.59$, and $\sigma = 0.1$.

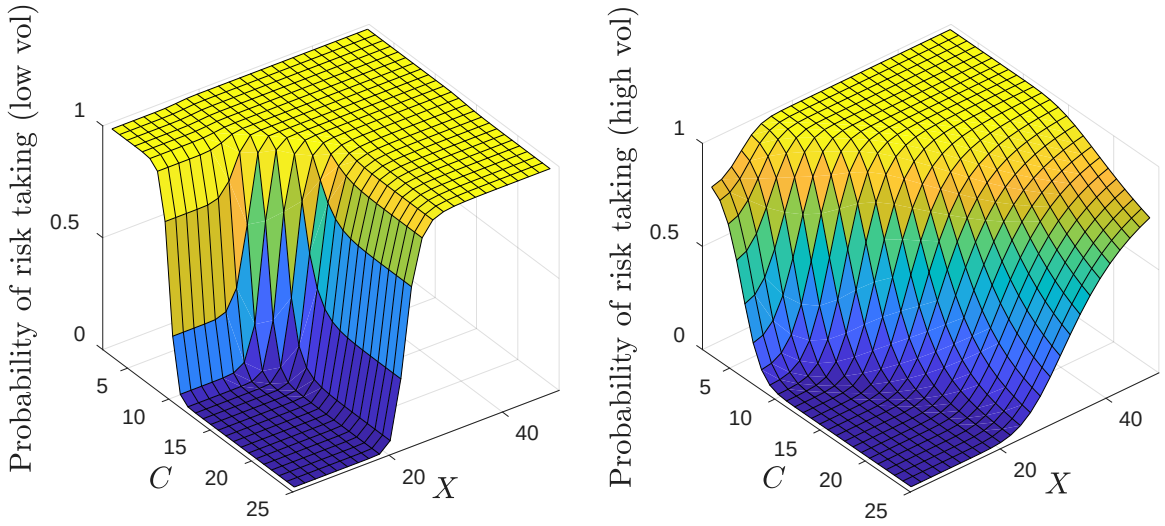


**Figure 4. The probability of risk taking as a function of *X* and *C*.** The figure plots, for two different volatility levels of the stimulus distribution, $\sigma_x = \sigma_c = 0.19$ (low volatility) and $\sigma_x = \sigma_c = 0.55$ (high volatility), the probability of risk taking computed in (18) as a function of both $X$ and $C$. The parameter values are: $\mu_x = 3.05$, $\mu_c = 2.35$, $p = 0.59$, and $\sigma = 0.1$.

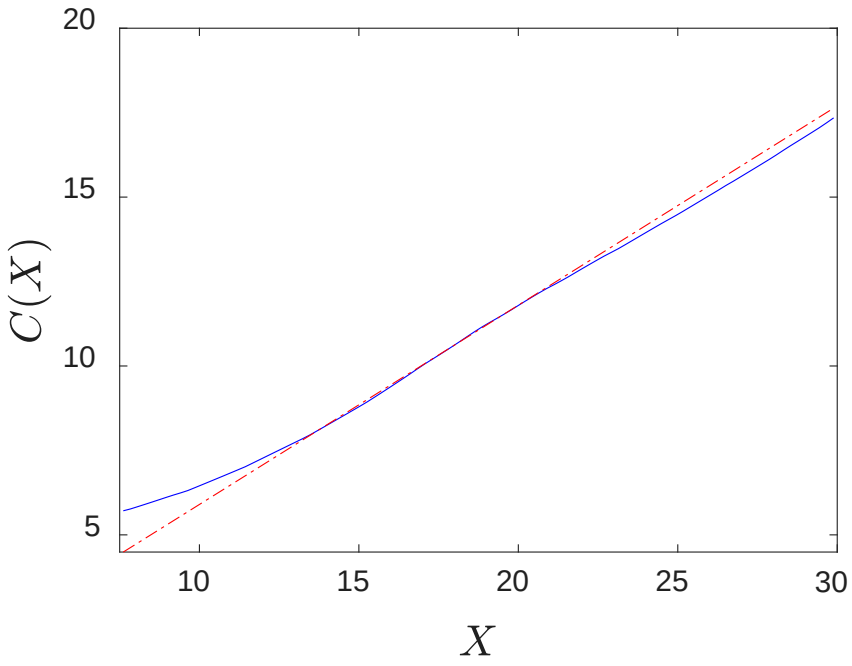


**Figure 5. Indifference curves.** Each curve plots the values $C(X)$ for which the *DM* is indifferent between the risky lottery and the certain option (that is, when $\mathbb{P}\text{rob}(\textit{risk taking} \mid X, C(X)) = 0.5$). The red dotted line plots the values for a *DM* that exhibits noiseless perception ($\sigma = 0$). The blue line plots the values for a *DM* who exhibits noisy perception ($\sigma = 0.1$). The parameter values are: $\mu_x = 3.05$, $\mu_c = 2.35$, $p = 0.59$, $\sigma_x = 0.55$, and $\sigma_c = 0.55$.

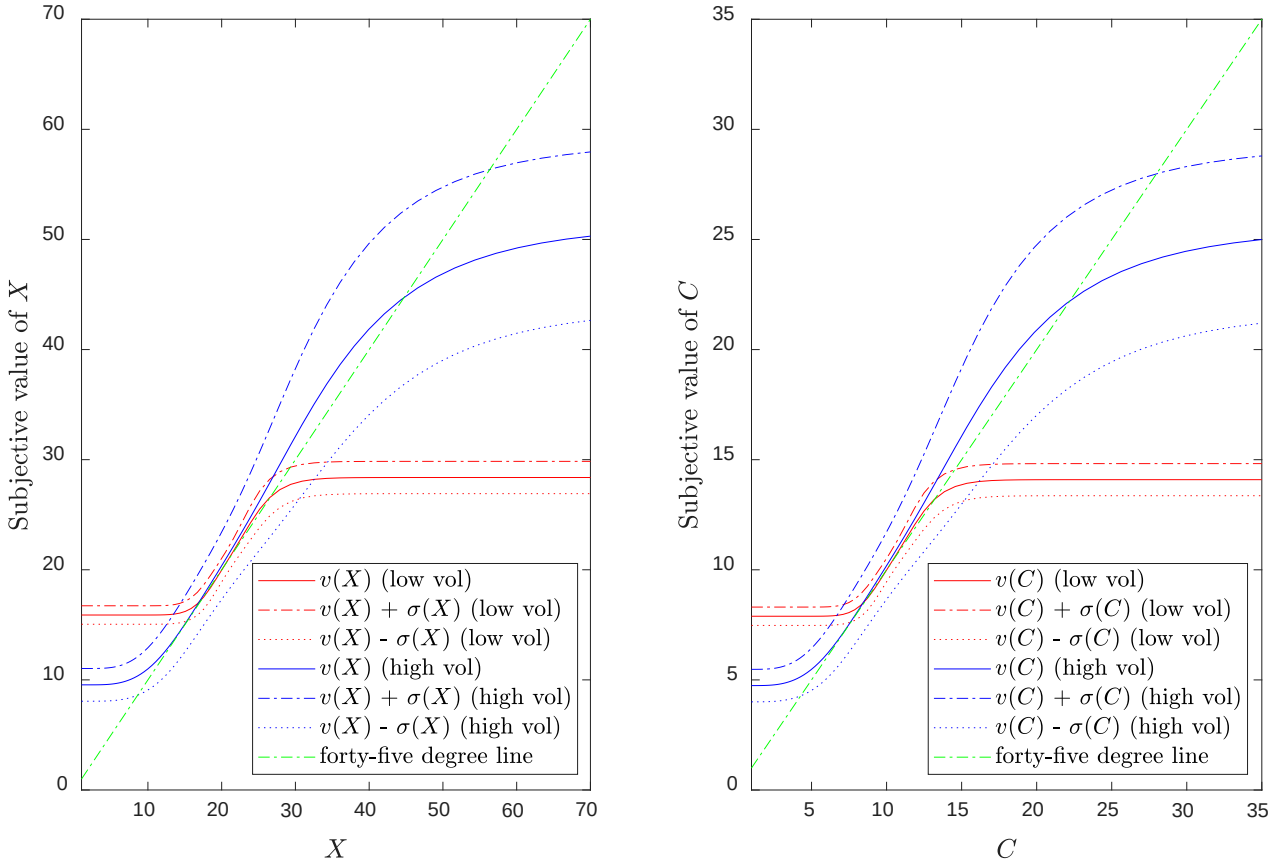


**Figure 6. Subjective value function shifts with the stimulus distribution.** The figure plots, for $X$ (left) and $C$ (right), and for both $\sigma_x = \sigma_c = 0.19$ (low-volatility condition, in red) and $\sigma_x = \sigma_c = 0.55$ (high-volatility condition, in blue), the subjective values $v(X)$ and $v(C)$, their one-standard-deviation bounds $v(X) \pm \sigma(X)$ and $v(C) \pm \sigma(C)$, as well as the forty-five degree line. The parameter values are: $\mu_x = 3.05$, $\mu_c = 2.35$, and $\sigma = 0.1$.

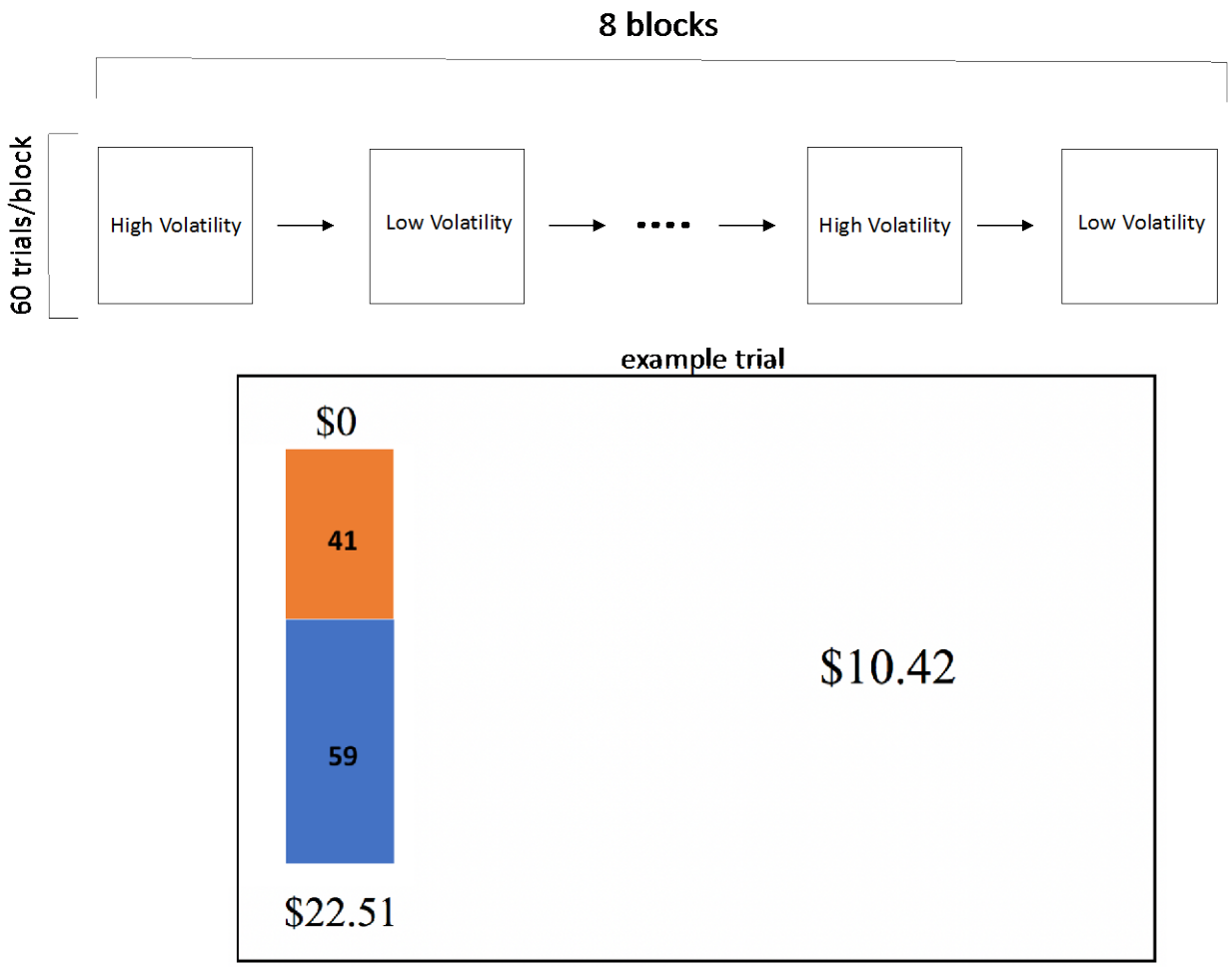


**Figure 7. Experimental design for risky choice task.** The task consists of eight blocks, which alternated between a high volatility condition and a low volatility condition. In the example trial screenshot above, the risky lottery is shown on the left, and the certain option is shown on the right. In each trial, the subject has unlimited time to decide which of the two options she prefers. After completing each block, the subject is allowed to take a self-paced break, after which the next block begins.

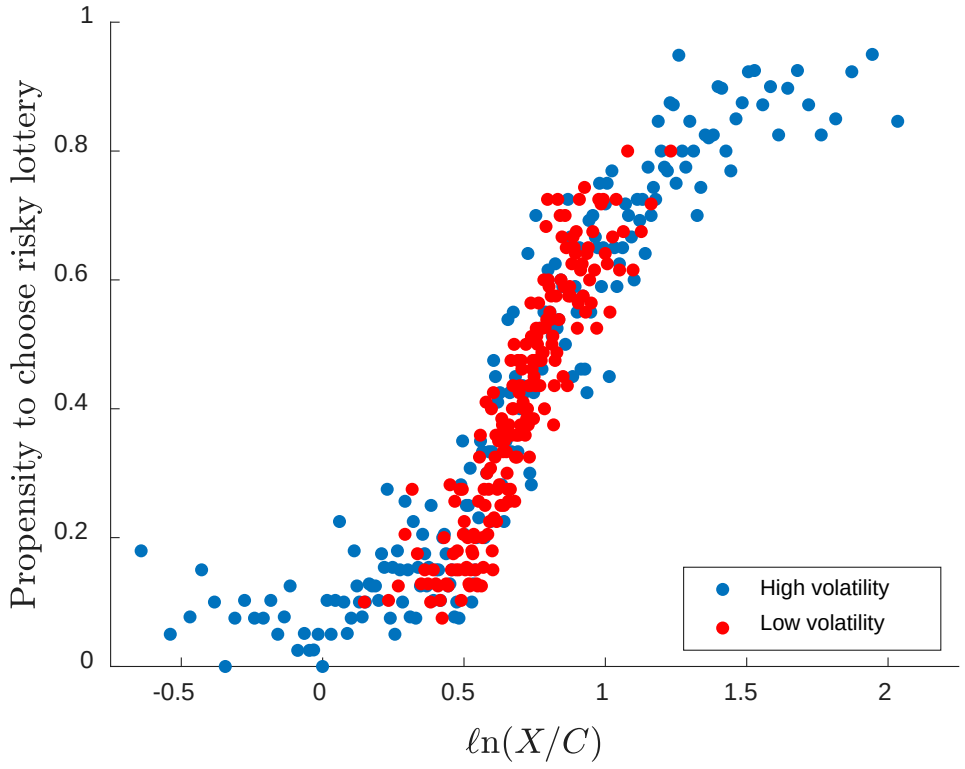


**Figure 8. Average levels of risk taking across conditions**. Data are pooled across trials and subjects. For each of the two experimental conditions, we bin the $\ell n(X/C)$ variable into two-hundred bins such that each bin has an equal number of trials.

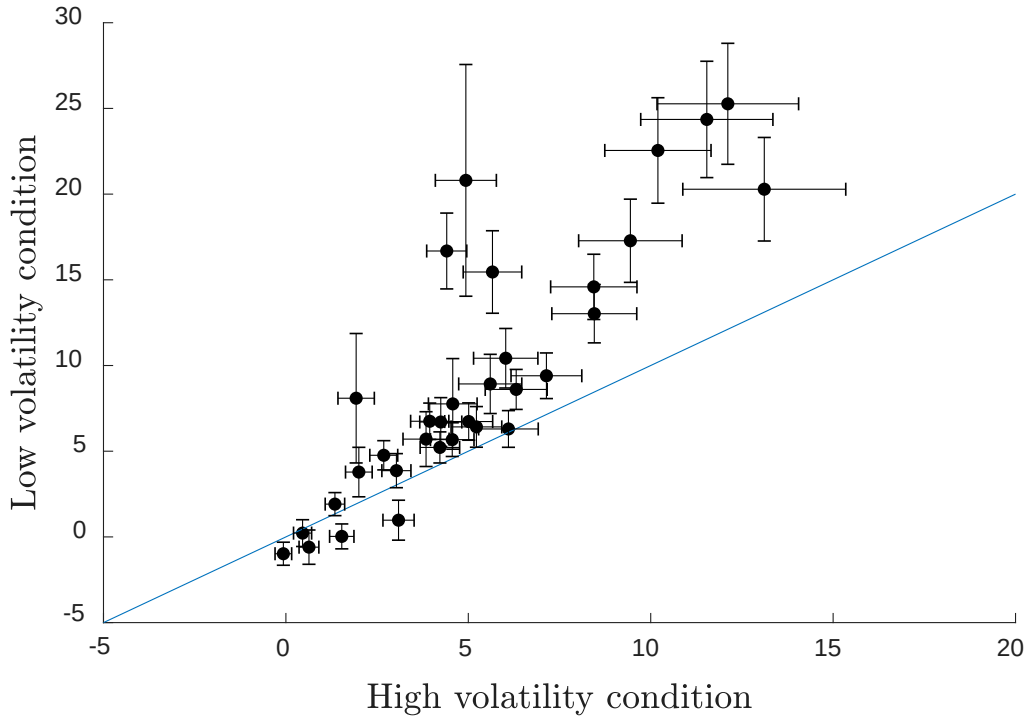


**Figure 9. Individual subject estimated treatment effects.** For each subject, and each condition, we run a logistic regression of the form: $risky_t = \alpha + \beta \cdot \ell n(X_t/C_t) + \varepsilon_t$. The *x*-axis measures the estimated β in the high volatility condition, while the *y*-axis measures the estimated β in the low volatility condition. Each point represents a single subject, and the length of each black bar denotes two standard errors of the mean. The blue line is the forty-five degree line.

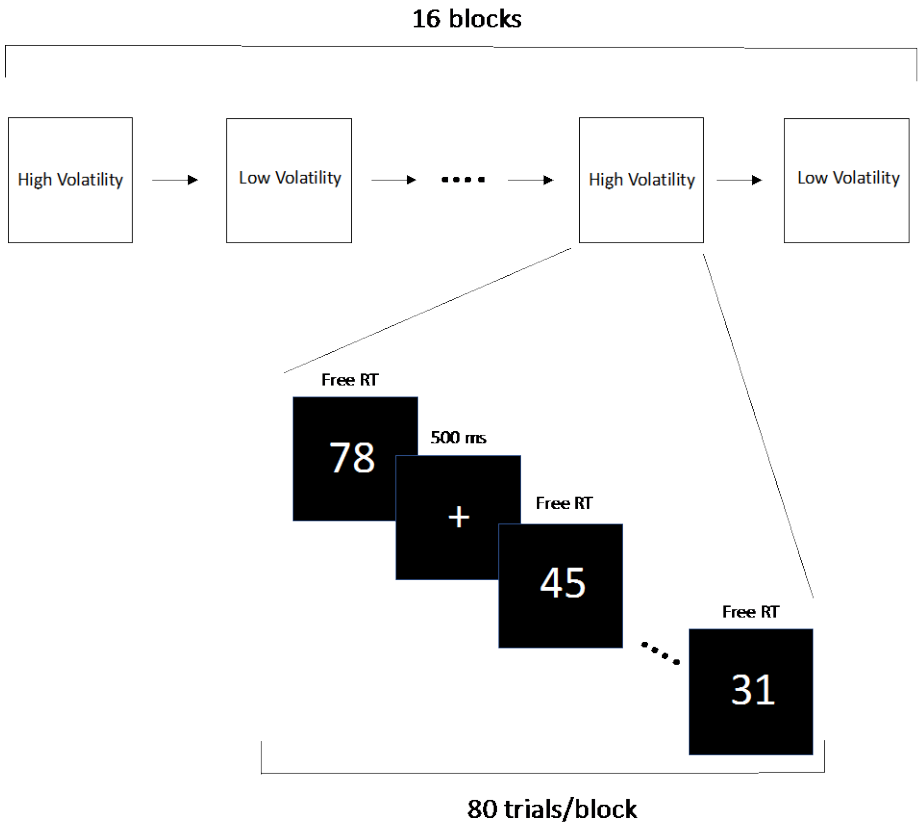


**Figure 10. Experimental design for riskless choice task.** The task consists of sixteen blocks, which alternated between a high volatility condition and a low volatility condition. On each trial, the subject is incentivized to classify as quickly and accurately as possible, whether the stimulus integer is larger or smaller than the number 65. In the high volatility condition, the integers are drawn uniformly from [31, 99], while in the low volatility condition, the integers are drawn uniformly from [51, 79].

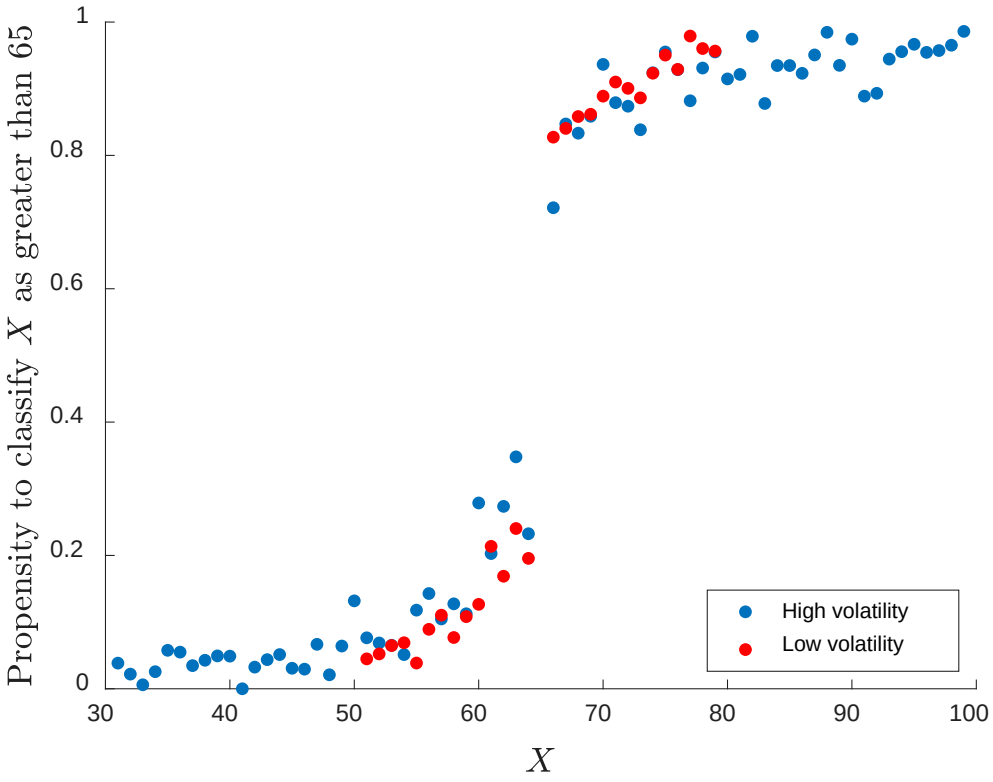


**Figure 11. Classification of numbers in riskless choice task.** The *x*-axis denotes the integer *X* that is presented on each trial. The *y*-axis denotes the proportion of trials for which the subject classified the integer *X* as greater than 65. Data are disaggregated by high and low volatility condition. In the high volatility condition, the integers are drawn uniformly from [31, 99], while in the low volatility condition, the integers are drawn uniformly from [51, 79].

## Panel A

| Dependent variable: "Choose risky lottery" | (1) All data | (2) $0.02 < \ell n(X/C) < 1.36$ | (3) First half of block | (4) Second half of block | (5) First 10 trials of each block | (6) First trial of each block |
|---|---|---|---|---|---|---|
| *high* | 0.99*** | 0.68*** | 0.84*** | 1.15*** | 1.10*** | 2.05** |
| | (0.28) | (0.26) | (0.30) | (0.32) | (0.29) | (1.04) |
| $\ell n(X/C)$ | 4.21*** | 4.21*** | 4.05*** | 4.37*** | 3.94*** | 5.05*** |
| | (0.62) | (0.62) | (0.62) | (0.65) | (0.70) | (1.27) |
| $\ell n(X/C) \times high$ | –1.35*** | –0.92*** | –1.11*** | –1.59*** | –1.14*** | –2.43** |
| | (0.38) | (0.32) | (0.40) | (0.42) | (0.41) | (1.22) |
| Constant | –3.38*** | –3.38*** | –3.26*** | –3.51*** | –3.26*** | –4.15** |
| | (0.47) | (0.47) | (0.47) | (0.50) | (0.50) | (0.96) |
| Pseudo $R$-squared | 0.17 | 0.13 | 0.18 | 0.17 | 0.17 | 0.19 |
| Observations | 15,840 | 14,101 | 7,920 | 7,920 | 2,640 | 264 |

## Panel B

| Dependent variable: "Choose risky lottery" | (1) All data | (2) First half of block | (3) Second half of block | (4) First 10 trials of each block | (5) First trial of each block |
|---|---|---|---|---|---|
| *high* | –0.10 | –0.20 | –0.01 | 0.03 | –0.42 |
| | (0.29) | (0.30) | (0.33) | (0.49) | (1.18) |
| $X$ | 0.18*** | 0.19*** | 0.17*** | 0.18*** | 0.21*** |
| | (0.03) | (0.03) | (0.03) | (0.03) | (0.07) |
| $C$ | –0.41*** | –0.42*** | –0.41*** | –0.39*** | –0.54*** |
| | (0.06) | (0.07) | (0.06) | (0.07) | (0.14) |
| $X \times high$ | –0.08*** | –0.09*** | –0.08*** | –0.07*** | –0.13** |
| | (0.02) | (0.02) | (0.02) | (0.03) | (0.06) |
| $C \times high$ | 0.18*** | 0.19*** | 0.16*** | 0.17*** | 0.34** |
| | (0.04) | (0.04) | (0.04) | (0.04) | (0.14) |
| Constant | 0.03 | –0.05 | 0.12 | –0.19 | 0.59 |
| | (0.42) | (0.43) | (0.45) | (0.53) | (1.19) |
| Pseudo $R$-squared | 0.15 | 0.15 | 0.15 | 0.15 | 0.16 |
| Observations | 15,840 | 7,920 | 7,920 | 2,640 | 264 |

**Table 1. Logistic regressions of probability of risk taking.** In both panels A and B, the dependent variable takes the value of one if the subject chose the risky lottery, and zero if the subject chose the certain option. The dummy variable *high* takes the value of one if the trial belongs to the high volatility condition, and zero if it belongs to the low volatility condition. Standard errors are clustered at the subject level, and ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

| Dependent variable: "Choose risky lottery" | (1) All data | (2) First and last 20 trials | (3) First and last 10 trials | (4) First and last 5 trials |
|---|---|---|---|---|
| $\ell n(X/C)$ | 3.07*** | 3.10*** | 3.41*** | 3.40*** |
| | (0.46) | (0.46) | (0.49) | (0.50) |
| $\ell n(X/C)\times outlier$ | –0.39** | –0.38** | –0.61*** | –0.70* |
| | (0.18) | (0.17) | (0.23) | (0.41) |
| $\ell n(X/C)\times second$ | –0.08 | –0.08 | –0.53** | –0.63*** |
| | (0.09) | (0.13) | (0.21) | (0.27) |
| $\ell n(X/C)\times second\times outlier$ | 0.15 | 0.02 | 0.57* | 1.14** |
| | (0.18) | (0.20) | (0.31) | (0.53) |
| Constant | –2.51*** | –2.47*** | –2.53*** | –2.43*** |
| | (0.37) | (0.36) | (0.35) | (0.38) |
| Pseudo $R$-squared | 0.25 | 0.25 | 0.27 | 0.27 |
| Observations | 5,940 | 3,960 | 1,980 | 990 |

**Table 2. Adaptation and sensitivity to outliers.** Logistic regression results using only data from the high volatility blocks (except for the first block, which does not have an immediately preceding low volatility block). The dependent variable *outlier* takes the value of one if the value of $\ell n(X/C)$ is more than three standard deviations from the mean, where these statistics are calculated using the sample moments from the sixty trials in the immediately preceding low volatility block. The dummy variable *second* takes the value of one if the trial belongs to the second half of the block (trials 31-60), and zero if it belongs to the first half of the block (trials 1-30). Standard errors are clustered at the subject level, and ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

| | (1) | (2) | (3) |
|---|---|---|---|
| Dependent variable: "Classifiy as greater than 65" | All data | $51 \leq X \leq 79$ | $X < 60$ or $X \geq 70$ |
| *high* | 0.00 | 0.04 | –0.07 |
| | (0.02) | (0.05) | (0.04) |
| $\ell$n($X$/65) | 19.53*** | 19.53*** | 17.45*** |
| | (2.09) | (2.09) | (1.58) |
| $\ell$n($X$/65)×*high* | –9.71*** | –3.56*** | –7.98*** |
| | (1.11) | (0.95) | (0.69) |
| Constant | 0.20*** | 0.20*** | 0.30*** |
| | (0.07) | (0.07) | (0.08) |
| Pseudo $R$-squared | 0.53 | 0.48 | 0.63 |
| Observations | 16,640 | 11,892 | 12,807 |

**Table 3. Classification in riskless choice task.** Logistic regression where the dependent variable takes the value of one if the subject classified the stimulus, $X$, as larger than 65, and zero otherwise. The dummy variable *high* takes the value of one if the trial belongs to the high volatility condition, and zero if it belongs to the low volatility condition. In the high volatility condition, the integer $X$ is drawn uniformly from [31, 99], while in the low volatility condition, the integer is drawn uniformly from [51, 79]. Standard errors are clustered at the subject level, and ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

# Appendix A: Theoretical Derivations

### A.1. Properties of $p(R_x)$

Given $p(X; \mu_x, \sigma_x)$ in (8) and $p(R_x \mid X)$ in (10), the distribution of $R_x$ can be derived as

$$\begin{aligned} p(R_x) &= \int_0^\infty p(R_x \mid X)\, p(X) dX \\ &= \frac{1}{2\pi\sigma} \int_0^\infty \exp\left( -\frac{(\Phi((\ell \mathrm{n}\, X - \mu_x)/\sigma_x) - R_x)^2}{2\sigma^2} \right) \frac{1}{\sigma_x X} \exp\left( -\frac{(\ell \mathrm{n}\, X - \mu_x)^2}{2\sigma_x^2} \right) dX \qquad \text{(A1)} \\ &= \frac{1}{2\pi\sigma} \int_0^\infty \exp\left( -\frac{(\Phi(y) - R_x)^2}{2\sigma^2} \right) \exp\left( -\frac{y^2}{2} \right) dy. \end{aligned}$$

Notice that this expression does not depend on distribution parameter $\mu_x$ and $\sigma_x$. Furthermore, this "invariance" result is a general statement independent of the specific assumption of lognormal distribution for $p(X)$: all continuous stimulus distributions lead to the same $p(R_x)$. To see this, we write in general

$$p(R_x \mid X) = g\left(\int_{-\infty}^{X} p(y) dy - R_x\right). \qquad \text{(A2)}$$

Then

$$\begin{aligned} p(R_x) &= \int_0^\infty p(R_x \mid X)\, p(X) dX \\ &= \int_0^\infty g\left(\int_{-\infty}^{X} p(y) dy - R_x\right) p(X) dX = \int_0^1 g(z - R_x) dz, \qquad \text{(A3)} \\ &\text{where } \; z \equiv \int_{-\infty}^{X} p(y) dy. \end{aligned}$$

This equation makes it clear that not only is the case that $p(R_x)$ does not depend on $\mu_x$ and $\sigma_x$, it does not depend on the entire *shape* of the distribution $p(X)$. A sufficient condition for this "invariance" result is that *i*) the likelihood function is location-independent in the sensory space (as we assume in equation (5) in the main text), and *ii*) the transformation function from stimulus space to the sensory space is the cumulative density function of the stimulus value.

Next, we look at the asymptotic behavior of $p(R_x)$ as $\sigma$ goes to zero. From (A3) we know that

$$p(R_x) = \int_0^1 g(z - R_x) dz \xrightarrow[\delta \to 0]{} \int_0^1 \delta(z - R_x) dz = \begin{cases} 1 & 0 < R_x < 1 \\ 0 & \text{otherwise} \end{cases}, \qquad \text{(A4)}$$

where $\delta(\cdot)$ represents the Dirac delta function. That is, in the limiting "noiseless" case, $p(R_x)$ converges to a uniform distribution between zero and one.

# Appendix B: Experimental Instructions

### B.1. Instructions for Risky Choice Task

Experiment Instructions

Thank you for participating in this experiment. Before we begin, please turn off all cell phones and put all belongings away. For your participation, you have already earned $7, and you will have the opportunity to earn more money depending on your answers during the experiment.

In the experiment, you will be asked to make a series of decisions about choosing a “risky gamble” or a “sure thing”. The risky gamble will pay a positive amount with 59% chance, and $0 with 41% chance. The amount shown for the sure thing will be paid with 100% chance, if chosen. Below is an example screen from the experiment:


In this example, the risky gamble pays $22.51 with 59% chance, and $0 with 41% chance. The sure thing pays $10.42 with 100% chance. You will be asked to select one of the two options for each question in the experiment. The experiment is broken down into eight parts, and each part contains sixty questions.

At the end of the experiment, one trial will be randomly selected, and you’ll be paid according to your decision on that trial. For example, if the above trial was chosen, and you selected the sure thing you would be paid a total of $10.42 + $7 = $17.42. If instead you chose the risky gamble, you’d be paid either $7 or ($22.51 + $7) = $29.51, depending on which outcome the computer randomly selects. Before we begin, you will see 10 practice trials to familiarize yourself with the software. These 10 practice trials will not count towards the real experiment.

**B.2. Instructions for Riskless Choice Task**

Experiment Instructions

Thank you for participating in this experiment. Before we begin, please turn off all cell phones and put away all belongings until the end of the experiment. For your participation, you have already earned $7, and you will have the opportunity to earn more money depending on your answers during the experiment.

In the experiment, you will see a series of numbers and will be asked to classify whether the number is larger or smaller than the number "65". If the number is larger than 65, press the "?" key, and if it is smaller than 65, press the "z" key. At the end of the experiment, you will be paid depending on the speed and accuracy of your classifications. Specifically, you will be paid:

$$\text{Payout} = \$(20 \times \textit{accuracy} - 10 \times \textit{avgseconds}),$$

where "*accuracy*" is the percentage of trials where you correctly classified the number as larger or smaller than 65. "*avgseconds*" is the average amount of time it takes you to classify a number throughout the experiment, in seconds. For example, if you correctly classified all trials and it took you 0.3 seconds to respond to each question, you would earn $\$(20 \times 100\% - 10 \times 0.3) = \$17.00$ (plus the $7 show-up fee). If instead you only answer 75% of the questions accurately and took 1 second to respond to each question, you would be paid $\$(20 \times 75\% - 10 \times 1) = \$5.00$ (plus the $7 show-up fee). Therefore, you will make the most money by answering as quickly and as accurately as possible.

The experiment will be separated into sixteen parts, and each part will contain 80 trials. In between each part, you can take a short (~1 minute) break, and then continue at your own pace. When you finish all sixteen parts, please raise your hand and do not disturb other subjects.

Before you begin the experiment, you will go through 10 practice trials to familiarize yourself with the software. These 10 practice trials will not be counted when computing your final payout.