Journal of Financial Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Outlier blindness: A neurobiological foundation for neglect of financial risk☆

Elise Payzan-LeNestour[a,*], Michael Woodford[b]

[a] UNSW Business School, University of New South Wales, Sydney, NSW, Australia
[b] Department of Economics, Columbia University, New York, NY, USA

ARTICLE INFO

Article history:
Received 14 July 2020
Revised 18 March 2021
Accepted 15 April 2021
Available online xxx

JEL classification:
C91
D87
G41

Keywords:
Tail risk
Imprecise perception
Efficient coding
Adaptation
Decision making under uncertainty
Behavioral finance
Neuroeconomics
Experiments

ABSTRACT

How do people record information about the outcomes they observe in their environment? Building on a well-established neuroscientific framework, we propose a model in which people are hampered in their perception of outcomes that they expect to seldom encounter. We provide experimental evidence for such "outlier blindness" and discuss how it provides a microfoundation for neglected tail risk by investors in financial markets.

1. Introduction

Most business activities involve making decisions with uncertain outcomes, and, even though predicting the future is impossible, keeping track of how often different things happen can help one to make a better decision. As (Dixit and Nalebuff, 1991) suggest, "Even though you can't guess right all the time, you can at least recognize the odds." However, behavioral scientists have long established that "recognizing the odds," i.e., learning outcome probabilities from observing the realized outcomes in one's environment, is itself no simple task.

Many discussions of this issue emphasize the inherent difficulty in telling apart different possible statistical models based on finite data samples, or many people's limited understanding of the principles of statistical inference. However, these discussions typically take for granted that

☆ We are indebted to Patrick Bolton, Peter Bossaerts, Brice Corgnet, Peter Dayan, Samuel Gershman, Philippe Gouillou, Lawrence Jin, Ronald Masulis, Carsten Murawski, Amos Nadler, Antonio Rangel, Christian Ruff, Philippe Tobler, seminar participants at Caltech, Melbourne University, the Behavioral and Experimental Economics and Finance workshop at the University of Sydney, the Department of Economics at the University of Zurich, University College London, and the 2019 American Finance Association, for very helpful comments and suggestions. Special thanks to Andrew Lo for discussing an early version of the paper at the 2019 American Finance Association, and to an anonymous reviewer for providing invaluable comments and suggestions. The analyses and conclusions set forth are those of the authors. Daniela Grinblatt, Philippe Gouillou, and Yeji Sung provided outstanding research assistance. Funding from Columbia University and UNSW Business School is gratefully acknowledged.

* Corresponding author.
E-mail address: elise@unsw.edu.au (E. Payzan-LeNestour).

https://doi.org/10.1016/j.jfineco.2021.06.019
0304-405X/

people are aware of exactly what has happened in the specific events that they have observed; it is the regularity that underlies the individual observations that is assumed to not always be learned. But, what if even perception, or the accuracy of recognizing and later recalling these observations, cannot be taken as given? In this study, we propose that people are hampered in their perception of outcomes that they expect to seldom encounter, and view the magnitude of such outcomes as less extreme than they really are. Metaphorically, it is as if people are "blind to outliers."

"Occam's Razor" might suggest that we should stick with the default assumption that perception is unbiased (on the ground that theories that proceed on this basis are simpler). But a recent neuroscientific paradigm called "efficient coding" suggests otherwise, presenting evidence that the brain is designed to communicate information in a way that economizes on its limited resources.[1] Since an equal focus on all possible values of a given stimulus would imply poor discrimination among different values, it seems plausible that the brain learns which outcomes are more likely to occur and allocates most of its resources to representing the outcomes frequently encountered in the environment at the expense of the unlikely outcomes. This could result in a diminished capacity to discriminate between outcomes far from the most likely ones ("outlier blindness").

While the principle of efficient coding has been extensively discussed in the neuroscience literature, the applicability of these ideas to financial decision-making may not be obvious. Here we offer both a novel theoretical exposition and new empirical evidence, to clarify the relevance of the efficient coding principle for finance. We propose a simpler mathematical formulation of the resource constraint that captures the essential structure of important models in the neuroscience literature, with the goal of providing a model that should be "portable" across domains in the sense advocated by Rabin (2013). We also analyze efficient coding under novel assumptions about the objective to be maximized and the statistics of the variables that are to be encoded, which we believe are more appropriate for applications to financial decision-making.

In our exposition, we emphasize in particular the result of outlier blindness, and an important implication of this general principle for financial decision-making, namely the neglect of tail risk. In our model, outlier blindness implies that we should expect to see underestimation of tail risk in financial investing, not because extreme events are not included in the sample on the basis of which risks are estimated, but because they are initially perceived to be less extreme than they actually are. For example, traders could initially underappreciate the size of volatility "jumps." Likewise, the importance of macroeconomic shocks such as the Long-Term Capital Management and Global Financial Crises could have initially been underappreciated.

How much of a problem such outlier blindness causes should depend on the statistics of the decision maker's environment. In our model, the distorted perception of outliers occurs even if the investing environment is frequently shifting (we elaborate in Section 3.1.2). This distorted perception should be particularly extreme in the case of distributions with especially long tails, which unfortunately is a common feature in financial markets.[2] Our model further predicts quick investor adaptation to a shift in the return distributions, such that outlier blindness occurs even if the investing environment is frequently changing. This contrasts with recent theories that imply that investor adaptation should be quite slow (e.g., Robson and Whitehead, 2019).

We provide experimental evidence for our model in a perceptual task specifically designed to measure outlier blindness in the laboratory. Essentially, we compare how well task participants discriminate between two given values when these values are within the range of values to which they have been exposed ("adapted") in the previous trials, versus when these values are outliers (outside the range to which the task participants have been adapted). We find that participant perception of outliers is significantly hampered, consistent with our model. The experimental data further support the model prediction that investor adaptation takes only a small number of trials and hence, outlier blindness can occur even when financial returns are affected by changing economic conditions.

The phenomenon of outlier blindness could be an important factor in explaining the vulnerability of capitalist economies to financial instability. For example, the crisis triggered by the failure of Lehman Brothers in the fall of 2008 had much more severe consequences, owing to many institutions having taken positions that left them exposed to significant risk in such an eventuality. Many observers have been puzzled by the lack of a reaction by investors to the news of the housing bubble bursting about two years earlier. Despite evidence of a regime shift, investors continued to neglect tail risk (even though, from the summer of 2007, daily realized volatility began to rise above 20%, with peaks above 40%).[3] One explanation for this apparent paradox is that investors chose to ignore the 2006 regime shift news as a result of greed (Lo, 2017) or for other psychological reasons related to "motivated beliefs" (e.g., Cheng et al., 2014 and Benabou, 2015). The present study offers an additional, complementary explanation: the idea that investors underestimated the importance of the news as a consequence of outlier blindness. For a more recent example, Giglio et al. (2020) report that investor long-term market expectations did not adjust to the implications of the Covid-19 pandemic, which could be related to outlier blindness as well.

Outlier blindness theory also provides a neurobiological foundation for the neglect of financial risk to become particularly widespread after two decades of a "Great Moderation." Minsky (1986) has argued that investors be-

[1] For example, Laughlin (1981); Tobler et al. (2005); Wei and Stocker (2015), and Ganguli and Simoncelli (2016).

[2] For evidence on the fat-tailedness of asset returns, see, e.g., Mandelbrot (1957); Fama (1965); Gabaix et al. (2003, 2006), and Kelly and Jiang (2014).

[3] See Lo (2017) and Gennaioli and Shleifer (2018) for further discussion.

come progressively more inclined to underestimate tail risk the longer a period of relative macroeconomic and financial stability continues, so that periods of stability predictably give rise to the fragility that allows a crisis to occur. The current study provides a cognitive mechanism through which this can occur: in our model, after a prolonged period of good returns and low volatility, when investors imagine—simulate in their mind—the occurrence of extreme negative returns, they underestimate their size. As a result, they give such outcomes insufficient weight in decision making.[4]

Relation to the literature. This paper adds to a recent and growing literature on risk perception. The literature on "neglected risks" shows how agents can underestimate the probability of extreme events until they occur, as a result of a number of cognitive biases such as the availability and representativeness heuristics,[5] undersampling extreme events (Hertwig et al., 2004), and wrongly assuming a Gaussian functional form when learning about fat-tailed payoffs under model uncertainty.[6] Here we propose an additional source of tail risk neglect that is not learning-related. Rather, it pertains to the way that outliers are perceived as less extreme than they are in reality in the first instance (even before the process of inferring patterns from individual instances), leading to underweighting rare events.

One could wonder how such underweighting of rare events can be reconciled with the *overweighting* of low-probability events predicted by Prospect Theory, for which laboratory experiments have provided ample evidence. In addition, a wide range of empirical findings in finance (e.g., Barberis and Huang, 2008, Kumar, 2009, and Kumar et al., 2011), insurance (e.g., Barseghyan et al., 2013), and gambling (e.g., Barberis, 2011), suggest that people overweight rare events in their decision-making. While at first glance, this could seem to be at odds with our theory, it need not be for two reasons.

First, and as noted by Barberis (2013), many of the foregoing applications of probability weighting studied in prior work can be thought of as "*decisions from description*," that is, settings in which the decision maker knows both the possible outcomes of a gamble and the probabilities with which each of them should occur because they are informed of these probabilities (as in many laboratory experiments), the probabilities can be deduced through simple reasoning, or historical data have been tabulated in a way that makes the (empirical) frequency distribution evident.

As we discuss further below, our theory of outlier blindness applies primarily to a different kind of setting, namely, "*decisions from experience*," in which people have to learn the distribution of the gambles by sampling from them, as is often the case in the business world. For instance, investment valuations,both in publicly traded securities and other asset classes such as private equity, and, of course, trading decisions, often rely on experience. Laboratory evidence suggests that, rather than being overweighted, low-probability outcomes are underweighted in the case of decisions from experience (e.g., Hertwig et al., 2004, Ungemach et al., 2009, and Erev et al., 2010). However, the mechanisms underlying such underweighting are not well-understood (e.g., Ungemach et al., 2009 and Barberis, 2013). Our results here suggest that outlier blindness could be among these mechanisms.

Furthermore, and importantly, the same principles of efficient coding that lead to our prediction of tail risk neglect in decisions from experience can also predict the well-established pattern of overweighting of rare events in decisions from description, when these principles are applied to the perception of outcome probability, a crucial aspect of decisions from description (see Section 3.3).

The paper also adds to a growing literature that seeks to understand the implications of "imprecise perception" [the idea that decisions are based on internal representations that are imprecise; see Woodford, 2020]. For example, Gabaix and Laibson (2017) show how imprecise representation of future payoffs can generate apparent temporal discounting in perfectly patient agents. Payzan-LeNestour et al. (2021) provide evidence for a perceptual bias ("after-effect") in investor risk perception. Frydman and Jin (2020) study how a change in the distribution of outcomes from one environment to another affects risk taking; they emphasize the effects of a change in the variance without changing the mean outcome. We instead consider the effects of a change in the mean of the distribution and focus on the phenomenon of outlier blindness. Robson and Whitehead (2019) also consider the effects of a change in the mean, but their focus differs from ours in that they emphasize the effects of a change in the coding rule on which stimuli should be considered "large" or "small," rather than on the fineness of the discriminations that are made between "outliers." Moreover, their study is purely theoretical.

The rest of the paper is organized as follows. Sections 2 and 3, respectively, describe the core prediction of our model and the experiments used to test it. Section 4 describes key extensions of the model and explains their implications for financial decision-making. Section 5 concludes by discussing several possible extensions of the current study for future behavioral finance research.

2. Benchmark model

2.1. From objective return magnitudes to their internal representations

To model investor perception of financial returns, we propose a new model that incorporates the basic structure of a number of models of efficient coding from the neuroscience literature, while abstracting from any specific biophysical model of how neurons represent environmental features. Moreover, and importantly, our model defines efficiency of a coding scheme in terms of the reliability of the decisions that are made on the basis of imprecise internal representations, rather than assuming that some measure of the information content of the internal representa-

[4] We thank Patrick Bolton for bringing to our attention this connection between Minsky's hypothesis and our theory of outlier blindness.

[5] For example, Gennaioli et al. (2012) and Gennaioli et al. (2015).

[6] For example, Taleb (2004); Donnelly and Embrechts (2010), and Payzan-LeNestour (2018).

tion is itself the criterion for efficiency, as is often assumed in the neuroscience literature.

We provide a detailed exposition of our model, and the ways in which it differs from previous expositions of efficient coding theory, in the Appendix. For brevity, this section and Section 3 only present the gist of the model.

We assume that when a given investor observes a return of magnitude x, the investor's brain produces an internal representation, the information that can be summarized by a sufficient statistic r (which is assumed to be a single number for simplicity). The representation r is drawn from a probability distribution conditional on the true return magnitude, which we denote by $p(r|x)$. If two return values x_1 and x_2 are presented and the investor must decide which one is greater, the investor's judgment must be based on the internal representations r_1 and r_2 evoked by the respective return values.[7]

In our baseline model in this section, we assume that mental processing is efficient in the sense of maximizing the probability of the investor correctly judging which of two magnitudes, x_1 and x_2, is larger, taking into account inherent limits on the precision of internal representations (which we further specify below). This objective corresponds to the reward structure in the experiment discussed in the next section. In Section 3, we show that the qualitative conclusions of our model still obtain under a variety of alternative choices for the efficiency criterion.

In the case of a binary comparison of the kind considered here, under fairly weak conditions, the investor should respond that the second return value is greater if and only if $r_2 > r_1$.[8] The probability of an erroneous judgment then depends on the degree of overlap of the two conditional distributions $p(r|x_1)$ and $p(r|x_2)$. When there is little overlap, as in the case shown in the upper panel of Fig. 1, in which the means of the two conditional distributions are far apart, judgments should almost always be correct. (Here we assume that the true values satisfy $x_2 > x_1$; thus, an error occurs if $r_2 < r_1$.) When instead there is substantial overlap, as in the case shown in the lower panel (the means of the two conditional distributions are close to each other), errors occur with substantial frequency (though still less than half the time).

Following a long-standing tradition in psychophysics (dating at least to Thurstone, 1927), we take the conditional distribution of values for r to be a normal distribution $N(m(x), \sigma^2)$, where $m(x)$ is a continuously increasing function of x, and σ is the same for all returns. The function $m(x)$ thus maps the objective return magnitude x to a so-called "Thurstone scale." The distance between two return magnitudes on the scale indicates the degree of overlap between the two conditional return distributions, that is, the extent to which the two returns are distinguishable (in the sense illustrated in Fig. 1). So, the steeper the function $m(x)$ over a given range of values of x, the sharper the investor's ability to discriminate between realized returns within this range.

To reflect the fact that the degree of precision with which different values can be distinguished is limited by the available resources of the brain (we have a finite number of neurons and each neuron can vary its activity over only a finite range), we assume that the function $m(x)$ takes values within some bounded interval $[\underline{m}, \bar{m}]$. The ratio $K \equiv |\bar{m} - \underline{m}|/\sigma$ determines the number of distinct return values that can be effectively distinguished from one another.

2.2. Outlier blindness

The basic idea at the core of efficient coding theory is that, given the limited available resources of the brain, increasing the number of states that can be effectively discriminated over one interval of the state space requires a corresponding reduction in the number that can be effectively discriminated elsewhere. To put it differently, there is an inherent trade-off between the degree of accuracy of perception in different parts of the state space. The optimal solution to this trade-off (i.e., the "efficient" way to allocate resources) will depend on one's objective; here we assume that the function $m(x)$ is chosen to maximize the probability of a correct response in our binary comparison task, in which the computation of this probability depends on a prior distribution over the comparisons that one might face.[9]

We show that to maximize the probability of a correct response in the foregoing binary comparison of return values, it is optimal to allocate fewer possible distinctions to regions of the return space that are infrequently encountered, given that the capacity to make fine distinctions in such cases will seldom matter. Instead, finer distinctions should be made in those parts of the return space where values are most frequently observed. This is accomplished computationally by having the $m(x)$ function increase most steeply in the part (or parts) of the return space with the greater prior density. Here we sketch the basic intuition, leaving the details of the analysis for the Appendix.

Let $F(x)$ be the cumulative distribution function (CDF) for the prior distribution from which returns are drawn in some environment; thus, $q = F(x)$ is the quantile of the return x. Any monotonic encoding rule $m(x)$ can then alternatively be specified as $m(x) = n(F(x))$, where $n(q)$ is a monotonic function that maps quantiles into the Thurstone scale. In our model, the probability of judging a quantity x_1 to be greater than another quantity x_2 depends only on the distance $m(x_1) - m(x_2)$ between the locations to which they are mapped on the Thurstone scale. Thus, it depends only on $n(q_1) - n(q_2)$, where q_1 and q_2 are the quantiles of these two returns. At the same time, whether such a judgment is correct or incorrect depends only on the relative order of the two quantiles; thus, we can define the probability of a correct decision on the basis of the func-

[7] See Woodford (2020) for a discussion of this approach to modeling human perception.

[8] This is the optimal decision rule to maximize the frequency of correct responses if the conditional probabilities satisfy the monotone likelihood ratio property (i.e., for any true magnitudes x, x' such that $x' > x$, the likelihood ratio $p(r|x')/p(r|x)$ is an increasing function of r), and if any pair of returns are equally likely to be presented in either order (i.e., under the prior probability distribution for the true values (x_1, x_2), (x, x') and (x', x) are equally likely, for any magnitudes x, x').

[9] We discuss alternatives to this assumption in Section 3.1 below.

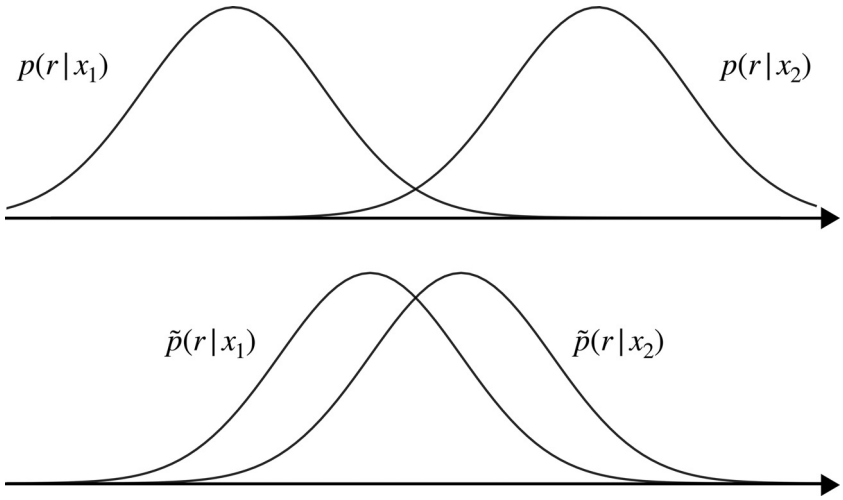

Fig. 1. Two examples of the distributions of possible internal representations of two return magnitudes (x_1 and x_2, with $x_2 > x_1$), with different implications for how distinguishable the return values are. Lower panel: the two return magnitudes are difficult to tell apart (there is considerable overlap between the two distributions). Upper panel: they are easily distinguished (there is little overlap).

tion $n(q)$ that maps quantiles to the Thurstone scale in a way that is independent of what the prior $F(x)$ could be.

It then follows that the quantile coding rule $n(q)$ that maximizes the probability of a correct decision (and hence that maximizes expected reward in our experiment) will be the same function $n^*(q)$, regardless of the prior distribution from which returns are drawn. The optimal encoding rule $m(x)$ will be different in the case of a different prior, but *only* because the assignment of quantile ranks q to given returns x has changed. Thus, if the prior $\tilde{F}(x)$ associated with a new environment is simply a mean shift of the prior $F(x)$ (so that $\tilde{F}(x) = F(x - \Delta)$ for all x, where Δ measures the size of the shift), and $m(x)$ is the optimal encoding rule for the original prior, then the optimal encoding rule for the new prior will be

$$\tilde{m}(x) \;=\; m(x - \Delta).$$

Hence, the degree of accuracy with which any two returns x_1 and x_2 can be distinguished is predicted to depend on the quantiles of these returns under the prior distribution. In the case of a prior under which both returns lie in a low-probability-density part of the distribution, their quantiles will not be very different, and they will be difficult to discriminate (the lower panel of Fig. 1).

We obtain an even stronger result for the limiting case in which the precision of internal representations is relatively high; in this case, we can show that the optimal quantile encoding rule $n^*(q)$ is close to linear. This means that the optimal encoding rule $m(x)$ satisfies

$$m'(x) \;\sim\; f(x), \tag{1}$$

where $f(x) = F'(x)$ is the probability density function implied by the prior. The steepness of the optimal encoding rule depends only on the prior density at that point.[10] In the case of a unimodal prior distribution, this implies that $m(x)$ should increase most steeply over an intermediate range near the mode of the frequency distribution of return magnitude x, and less steeply for more extreme values of x.

As an illustration of this key feature of the model, consider Fig. 2, in which the $m(x)$ function (bottom right panel) represents the efficient coding scheme in the case of an environment in which the return distribution is given by the normal distribution $f(x)$ (top right panel). Note that $m(x)$ increases most steeply at the value of x where $f(x)$ is highest, and progressively less steeply for more extreme values of x both above and below this value. Instead, the $\tilde{m}(x)$ function on the bottom right panel represents an efficient coding scheme when the return distribution is given by $\tilde{f}(x)$ (top right panel): $\tilde{m}(x)$ increases most steeply at the value of x where $\tilde{f}(x)$ is highest, and progressively less steeply for more extreme values.

Functions $m(x)$ and $\tilde{m}(x)$ have the same range of possible values (the horizontal axis of the lower right panel), and the same bounded range of variation ($[\underline{m}, \bar{m}]$ on the vertical axis, which represents the Thurstone scale mentioned above). Nonetheless, the degree to which the two magnitudes x_1 and x_2 are discriminated is different in the two cases, as shown by the differing degrees of overlap of the conditional distribution of values for the internal representation r (shown in the lower left panel of the figure). With the function $m(x)$, magnitudes x_1 and x_2 are well discriminated, whereas with $\tilde{m}(x)$, they should be frequently confused. The two alternative coding rules represent alternative uses of a fixed range of possible internal representations; which one is better will depend on the environment in which it is used. Return values like x_1 and x_2 are encountered more often in the case of distribution $f(x)$; hence, it is more important to be able to distinguish them well in that case, than if the environment is instead described by frequency distribution $\tilde{f}(x)$. In the latter case, the ability to tell apart x_1 and x_2 is sacrificed, for the sake of sharper discriminations in a different part of

[10] A similar result is obtained in a number of models of efficient neural coding, though under assumptions different from the ones that we make here, and less obviously appropriate for financial applications, as discussed in Appendix A.4.

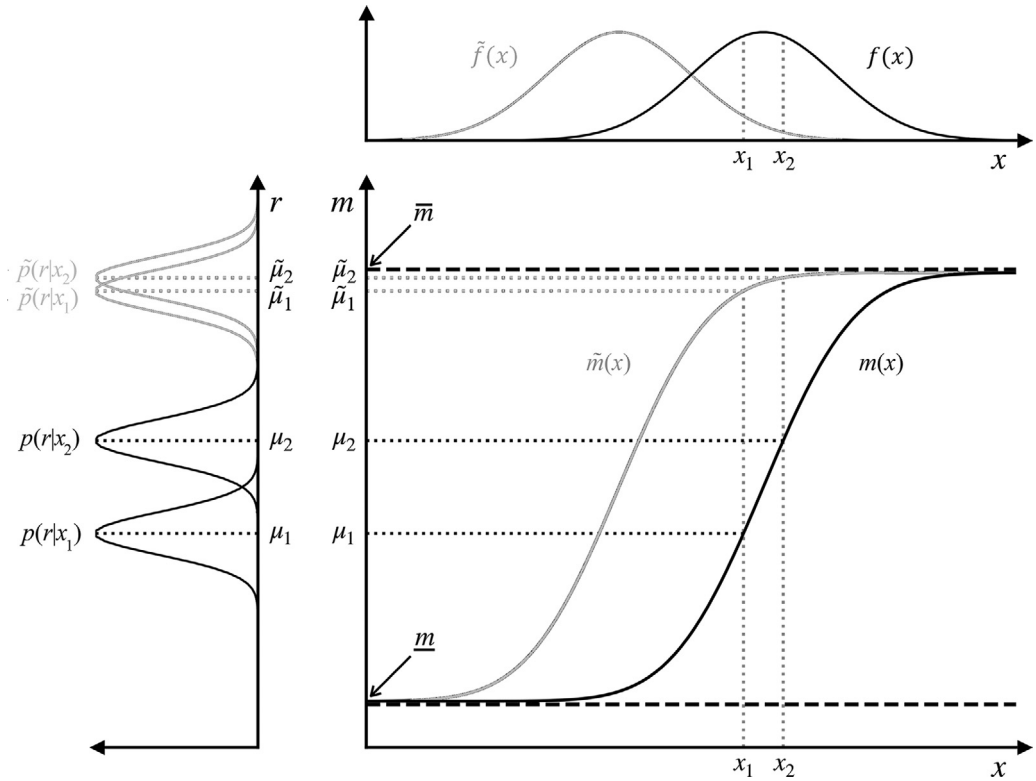

Fig. 2. Two possible distributions for the return magnitude x ($f(x)$ and $\tilde{f}(x)$, top right panel), and their efficient coding schemes (respectively, $m(x)$ and $\tilde{m}(x)$, bottom right panel). The degree to which the two magnitudes x_1 and x_2 are discriminated is different in the two cases, as shown by the differing degrees of overlap of the conditional distributions of values for their internal representations (lower left panel).

the return space (the part where the prior density $\tilde{f}(x)$ is larger).

It is clear from Fig. 2 that the $m(x)$ (resp. $\tilde{m}(x)$) efficient coding function is essentially flat in the tails of the frequency distribution $f(x)$ (resp. $\tilde{f}(x)$), implying that outliers (whether values in the far left tail or the far right tail) are poorly distinguished from one another. This is the phenomenon of outlier blindness mentioned in the introduction, which is our main focus in this study. In Section 3.2, we show that it leads people to underestimate the magnitude of tail events.

Of note, the phenomenon of outlier blindness should occur with any frequency distribution, as long as there is a long tail on at least one side of the distribution, with increasingly more extreme values occurring with lower and lower probability. The specific Gaussian functional form for the frequency distribution of the return magnitude x used in Fig. 2 (and our experimental task uses a Gaussian distribution as well; see Section 3) is not essential for outlier blindness to occur.

Fig. 2 further illustrates the important observation that what counts as an outlier depends on what one has reason to expect. The same two return values x_1 and x_2 are not outliers in the environment in which the frequency distribution of returns is described by $f(x)$, and hence they are fairly accurately discriminated; yet, the same realized returns are outliers in the other environment, in which the frequency distribution is instead described by $\tilde{f}(x)$, and hence they are frequently confused with one another.

This has important consequences for applications of the principle of outlier blindness to economic and financial decision-making (Section 3). One might well ask how one can say that there is lower discriminability of nearby states in one part of the state space than in another, if it isn't clear how "near" two states should be considered to be to one another, in a way that makes it possible to say that pairs of states in different parts of the state space are "equally distant" from one another. The problem is solved if we consider how the discriminability of *the same two states* changes when we change the frequency distribution of states for the context in which they occur. This provides the basis for a particularly compelling test of the existence of outlier blindness, which we report next.

3. Experimental test of the model

3.1. Experimental design

It would be difficult to conclusively demonstrate the existence of outlier blindness in investor perception of financial returns by observing investor choices in the field, owing to the inevitable presence of many confounding factors. Therefore, to test our model, we turn to controlled experimentation in a laboratory setting. Even in the case of a laboratory experiment, an experimental task in-

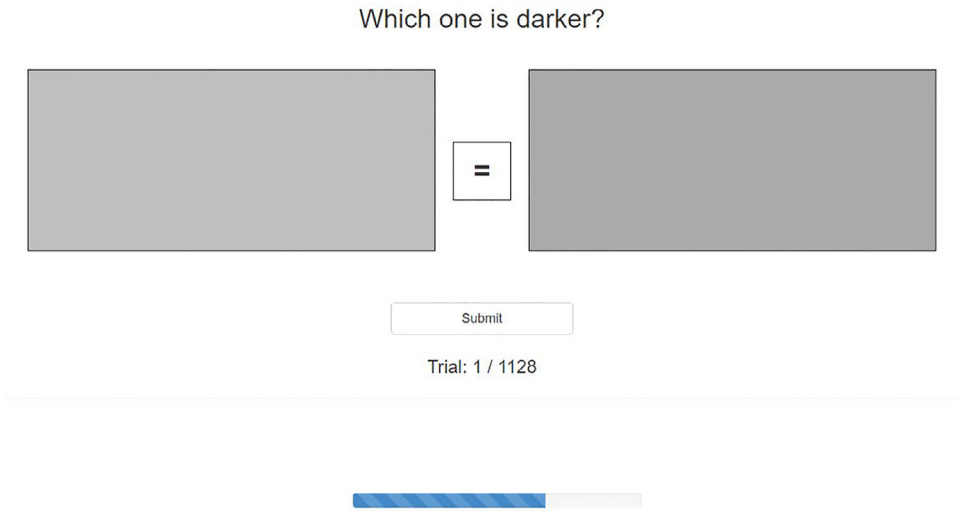

Fig. 3. User interface of the experimental task. On each of 1228 trials, the participant is to click on the rectangle that looks darker, or on an "=" icon displayed in the middle of the screen if the rectangles look to be the same color. The time allowed to make a reply is two seconds (remaining time is indicated through a timer at the bottom of the screen). One run of the task lasts for about 25 min.

volving economic payoffs would make a clear demonstration of our predicted effect difficult, as participant choices would reflect both their beliefs and their risk preferences, without our being able to disentangle these two determinants of their decisions. Hence, to be able to test the prediction of our model depicted in Fig. 2 above, we design a purely perceptual task in which on each trial, the participants have to discriminate between shades of grey in two rectangles (Fig. 3). This enables us to examine participants' perception per se, since any dependence on personal risk preferences is absent from the task by design.

One could wonder, of course, whether perception works in the same way for a low-level stimulus such as shades of grey as it does for the quantitative information on which economic decisions are based (financial returns, monetary payoffs, etc.). A long tradition in behavioral economics argues that there are important analogies between distortions in sensory perception (such as visual illusions) and distortions in the assessment of economic or financial opportunities, starting from (Kahneman, 2003). There is also a growing consensus in neuroscience that such analogies are pervasive.[11] In Section 4, we further discuss this point along with other questions also related to the applicability of the current findings to financial decision-making.

To test the prediction of our model depicted in Fig. 2 above, we compute participant accuracy (the fraction of correct replies) in "test trials" in which the shades that are presented to the participant are outliers in the sense that they are outside the range of shade values to which the participant has been exposed ("adapted") in the previous 40 trials. We then compare it to participant accuracy in "control trials" in which the exact same shades are presented to the participant, but this time they are not outliers as they are within the range of values to which the participant has been exposed in the previous 40 trials. See Fig. 4.

For example, a participant could be asked to discriminate between shade values in the range 2–5[12] for 40 trials, and then be asked to discriminate between shades 8 and 9 (test trial). At some other time during the task, the participant is going to be asked to discriminate between shades 8 and 9 again, this time after exposure to shades around 8 and 9 in the previous 40 trials (control trial). For each participant, there are 24 test/control trial pairs of this kind. The main statistic of interest is the difference between participant accuracy in the control vs. test trials: the model predicts a decreased accuracy for outlier perception, i.e., in the test trials, vis-à-vis the control trials.

All the information needed to replicate the following experimental findings is provided in the Online Appendix. We provide a detailed description of the experimental task (its stochastic structure in particular), the instructions for the task, and a detailed justification for each feature of the experimental design. The code to generate the experimental task used in each experiment, as well as the code to run the analyses reported in this section, can be downloaded at https://supplementarymat.weebly.com. The experimental data for each experiment are also provided there.

3.2. Results

3.2.1. Main findings

Sixty-three undergraduate students from the University of New South Wales participate in the experiment. A priori

[11] For example, Carandini and Heeger (2012); Glimcher (2014); Padoa-Schioppa and Rustichini (2014); Summerfield and Tsetsos (2015), and Bhui and Gershman (2018). Empirical evidence for this idea (that the principles underlying the perception of quantitative information and low-level stimuli are the same) can be found in Tremblay and Schultz (1999); Padoa-Schioppa and Assad (2008), and Khaw et al. (2017), among others.

[12] The scale of grey used in our task goes from 1 (very light) to 12 (very dark). See the Online Appendix for more details.

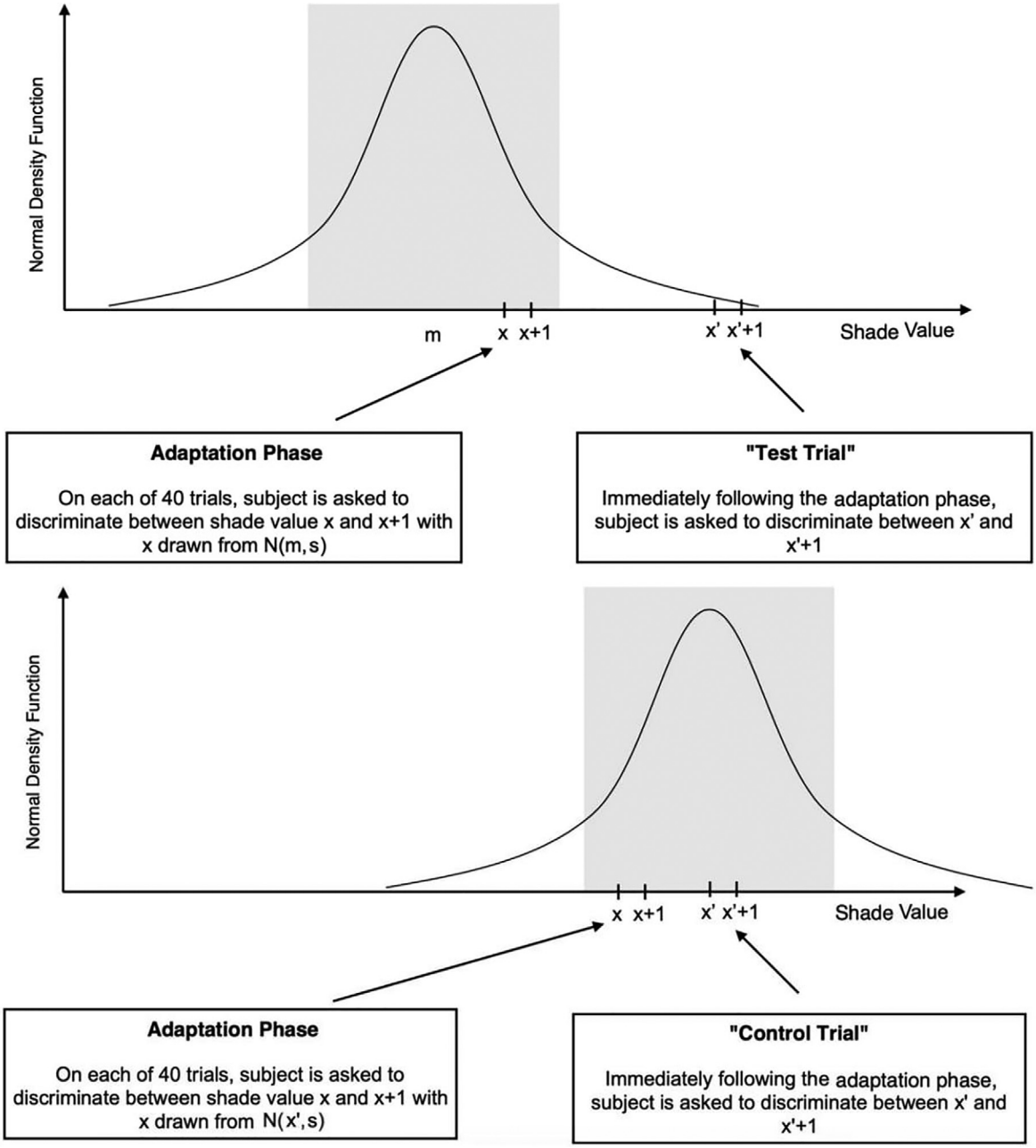

Fig. 4. Diagram explaining the core of the experimental strategy. Top graph: The participant goes through an adaptation phase in which on each of 40 trials, shade value x is drawn from a normal distribution with mean m and standard deviation s, immediately followed by a "test trial" with shade x' randomly drawn in the range of values located at least three standard deviations from m. Bottom graph: Each value x' in turn defines a 40-trial adaptation phase in which, on each trial, the shade value presented to the participant is normally distributed around x' (standard deviation is either 1 or 2; choice is random), followed by a "control trial" in which the shade values presented to the participant are exactly like those in the test trial (in this example, x' and $x'+1$). For each participant, there are 24 pairs of test and control trials of this kind. Values for m and s are randomly drawn in the sets $\{3, 4, \ldots, 10\}$ and $\{1, 2\}$, respectively.

power calculations conducted using G*power (Faul et al., 2007) show that this sample size gives us 98% power to detect a medium outlier blindness effect. To maximize data quality by ensuring that the participants pay attention on each trial, we provide them with high monetary incentives. One key aspect of this strategy consists of rewarding high performance in the task through high payoffs (for the other aspects, see the Online Appendix). Eight participants earn more than $70, 32 earn more than $50 (mean: 43.8; median: 50.10; mode: 45.65; std: 36.7). All aspects of our experimental protocol are approved by the UNSW Human Research Advisory Panel (HREAP) as part of the re-

Table 1
Tests of the outlier blindness effect, for each experiment. The outlier blindness effect is measured by the differential accuracy in the control vs. test trials as explained in the main text. Numbers in parentheses: standard error of the mean (sem).

	Mean accuracy in		Two-tailed	Wilcoxon signed
	test trials	control trials	paired t-test	rank test
Main experiment	0.61	0.79	$t = 11.76$	$V = 1858.5$
($N = 63$)	(0.02)	(0.01)	$p \sim 0$	$p \sim 0$
5-trial adaptation experiment	0.72	0.75	$t = 4.25$	$V = 330$
($N = 31$)	(0.01)	(0.01)	$p < 0.001$	$p = 0.006$
Placebo experiment	0.73	0.73	$t = 1.12$	$V = 396$
($N = 35$)	(0.01)	(0.01)	$p = 0.272$	$p = 0.131$
Experiment with double response time	0.72	0.84	$t = 6.27$	$V = 390$
($N = 33$)	(0.02)	(0.01)	$p < 0.0001$	$p < 0.0001$

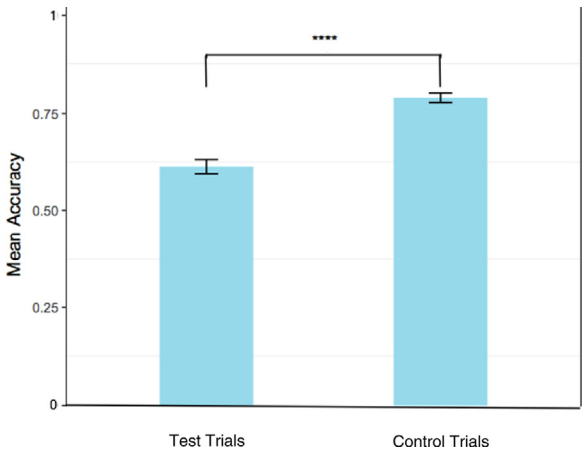

Fig. 5. Mean accuracy in the test trials (left) and control trials (right). Heights of the bars indicate the mean accuracy averaged across the 63 participants. Accuracy is defined as the fraction of correct replies. Line segments indicate standard error of the mean (sem). **** $p < 0.00001$. By design, the shade values presented in the test and control trials are the same; they are outliers when presented in the test trials and within the expected range when presented in the control trials.

quirements of the National Statement on Ethical Conduct in Human Research, as well as by UNSW Business School Experimental Research Laboratory for compliance with the standard rules imposed on all lab users. These rules include providing participants with significant monetary incentives, and the "no-deception rule."

The average response time in the experiment is 0.9 s (min: 0.66; max: 1.41; std: 0.14). There is a negative correlation between accuracy and response time (one-way Anova test: $f = 1454$; $p < 0.00001$, two-sided). Missed trials (trials in which the participant fails to reply within the required time) seldom occur (average percentage of missed trials across participants: 0.005%). Supplementary statistics (regarding response time, earnings, missed trials, and accuracy) are provided in the Online Appendix.

The main finding is that across participants, the perceptual accuracy for a given value is significantly lower when the value is an outlier (in the test trials) than when it is not (in the control trials), as predicted by our model. Our statistic of interest (i.e., the accuracy difference between control and test trials) averaged across participants is significantly positive (paired t-test: $t = 11.7$; $p \sim 0$; see Fig. 5 and Table 1). Strikingly, the statistic is positive for almost all the participants (60 out of 63; see Fig. 6).

3.2.2. Placebo test

To check that the foregoing findings truly reflect outlier blindness rather than other causes, we run a follow-up "placebo experiment," which replicates the design of the original experiment in all respects except that the participants ($N = 35$, same cohort as in the original experiment) do not have time to form expectations about shade values as the adaptation lasts for only three trials (the minimal achievable in the current design; see the Online Appendix for details). Thus, the values presented in the test trials do not count as "outliers" and hence there is no reason to expect decreased accuracy in the test trials in this case. Accordingly, we cannot reject the null hypothesis that accuracy is the same in the test and control trials of the placebo test; see Table 1. The absence of effect in the placebo test is also apparent in Fig. 7 which displays the distribution of accuracy in the test and control trials across participants, for both the original experiment (blue curves) and the placebo experiment (red curves). For the latter, the distributions of accuracy in the test and control trials overlap. This contrasts with the original experiment for which the outlier blindness effect is manifest (blue curves: the mean accuracy level at the test trials is shifted to the left relative to the mean accuracy at the control trials).

3.2.3. Evidence for high-speed adaptation

An important question for applications of our theory is how quick the brain is to adapt to changes in the frequency distribution of the magnitudes to be encoded. To test whether rapid adaptation is possible, we run the "five-trial adaptation experiment" ($N = 31$; same cohort as in the original experiment) in which adaptation lasts for only five trials. We find a significant outlier blindness effect in that experiment (paired t-test: $t = 4.2$, $p < 0.001$, two-tailed; see Table 1). Thus, an adaptation length of five trials appears to be sufficient for the agent to form expectations and hence for the outlier blindness effect to reappear.

Note that the effect found in our baseline experiment is only partially recovered: the outlier blindness effect is significantly larger in the original experiment than in the 5-trial adaptation experiment (two-sample t-test to compare the statistic of interest in the two experiments: $t = 8.8$,

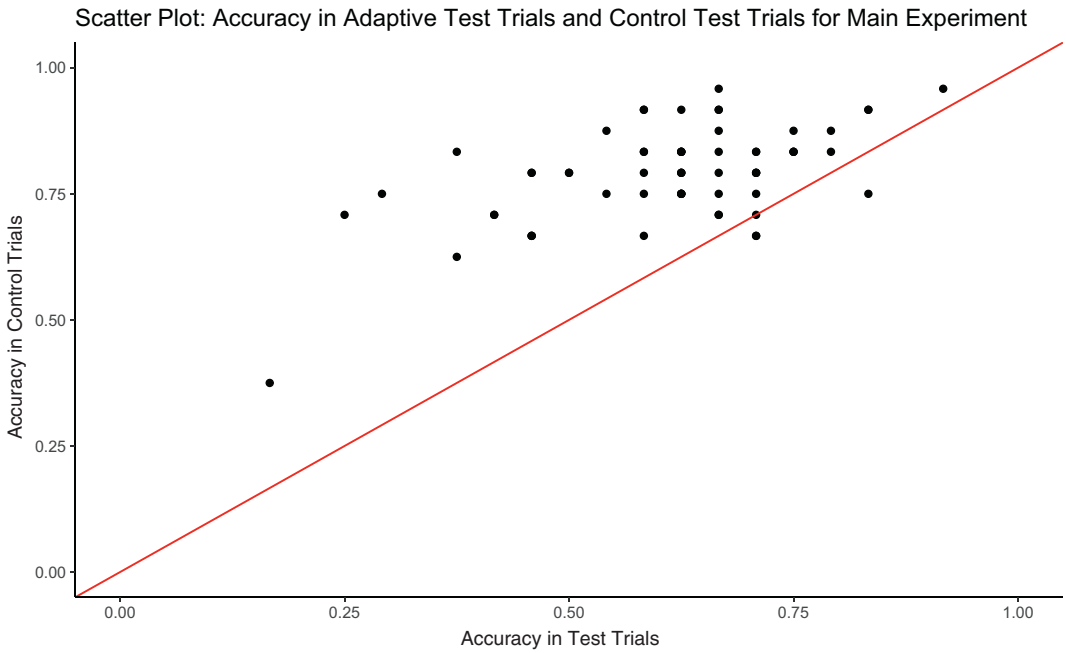

Fig. 6. Comparative accuracy levels in the test versus control trials for each task participant. Each data point corresponds to one participant ($N = 63$; 24 observations per participant). x axis: accuracy in the test trials. y axis: accuracy in the control trials. Data points above the 45-degree line correspond to participants for whom accuracy is decreased in the test trials, as predicted by our model.

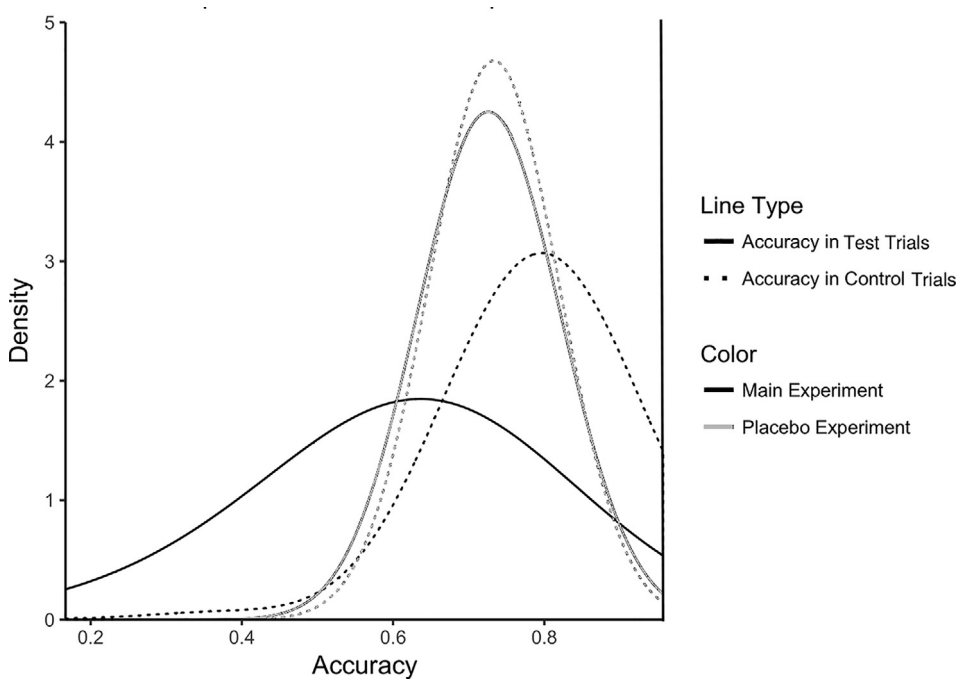

Fig. 7. Density plot of accuracy in the test and control trials, for both the original experiment (black) and the placebo test grey). The density is derived across participants based on 63 participants, 24 observations per participant (original experiment), and 35 participants, 320 observations per participant (placebo test).

Table 2
Comparison of the magnitude of the outlier blindness effect across experiments. In each experiment, the magnitude of outlier blindness is measured by the difference in participant accuracy in the control vs. test trials, averaged across participants. Numbers in parentheses: standard error of the mean (sem).

	(Mean accuracy in control trials) (Mean accuracy in test trials)	Two-sample t-test
Main experiment	0.18	$t = 1.94$, $p = 0.057$
($N = 63$)	(0.02)	
Experiment with double response time	0.13	
($N = 33$)	(0.02)	
Main experiment	0.18	$t = 8.89$, $p \sim 0$
($N = 63$)	(0.02)	
5-trial adaptation experiment	0.03	
($N = 31$)	(0.01)	

$p < 0.001$; see Table 2), as expected (a longer period of experience of a distribution of magnitudes allows more precise learning and hence more complete adaptation to the distribution).

3.2.4. Robustness checks

We run an extensive set of robustness checks, which include examining whether the outlier blindness effect is robust to an increasing response time by participants by running a final experiment in which the time allowed to provide a reply on each trial is doubled relative to that in the original experiment. We find a significant outlier blindness effect in that experiment (Table 1). We cannot reject the null hypothesis that outlier blindness is similar in magnitude in the original experiment and the experiment with double response time (Table 2).

4. Extensions of the model and implications for financial decision making

Having demonstrated the existence of outlier blindness in a setting in which the nature of the effect can be established unambiguously, we now study the generality of the outlier blindness result and key implications for financial decision-making.

4.1. Generality of the outlier blindness result

4.1.1. Consequences of assuming alternative objective functions

Some could argue that typical financial decisions involve judgments about magnitudes that are not even formally analogous to the problem considered in Sections 1 and 2, in which the objective is to correctly rank the relative magnitudes of two stimuli as frequently as possible. This is not obviously the right criterion in the case of financial decisions.

First, while it obviously is relevant to an investor to correctly recognize that the mean return on one asset is greater than that on another, it is not true that it should *only* matter to the investor how often she correctly ranks two expected returns; it is more valuable to correctly rank two returns when the difference is large than when it is small. An important question, therefore, is whether the result of outlier blindness still prevails if, instead of maximizing the fraction of correct rankings, the coding scheme is optimized to maximize the average magnitude of the return identified as larger by the decision maker. This alternative objective function is the one that would be relevant in the case of an investor who must choose which of two assets to invest in when the investor's reward is not a prize for making the "correct" decision (the reward scheme in a typical perceptual experiment), but rather the return on the investment that she chooses.

In Appendix A.5, we show that, in the case of this alternative criterion for efficiency, it remains the case that the optimal encoding function should increase more steeply over those parts of the return space with greater prior density. Specifically, we show that, because it is now particularly valuable to make correct judgments when larger quantities are at stake, it is no longer efficient to reduce discriminability in the tails of the distribution to quite the same extent as with the benchmark accuracy-of-ranking objective; but this effect is not, strong enough to overturn the result of outlier blindness. Recall that, in our baseline case, the optimal encoding rule satisfies (1). In the case of the alternative expected-value objective, this takes the more general form

$$m'(x) \sim f(x)^{\alpha}, \tag{2}$$

where now $\alpha = 2/3$, rather than having $\alpha = 1$ as with the accuracy-of-ranking objective. So, the slope of the encoding rule is still an increasing function of the prior density.

Another alternative criterion of relevance for some economic and financial decisions is that of minimizing the mean squared error of the estimated value $\hat{x}$ of the state, under the assumption that the estimate must be based on the internal representation r. To do well on this criterion, it is again necessary not simply to be able to recognize the correct ordinal ranking of states, but to be able to judge *how much* larger some are than others. Yet, again, we show that the optimal encoding rule under this alternative objective satisfies (2), except now with an exponent of $\alpha = 1/3$. So, the prediction of outlier blindness holds in this case as well.

Moreover, and importantly, our model accounts for the possibility that investors' losses from poor recognition of the magnitude of a monetary payoff may not be well described by any translation-invariant loss function of the

kinds considered above.[13] Instead, errors in recognizing the exact magnitude of the payoff might have more serious consequences for magnitudes in a certain range. For example, monetary losses larger than a particular size might become especially disastrous for an investor, so that the utility loss is a nonlinear function of the monetary loss. If such nonlinearities become particularly severe in the case of especially large (though unlikely) losses, the increasing significance for utility of further increases in the magnitude of the loss can be a reason for it to be valuable to make sharper discriminations between different large quantities, despite the infrequency with which they will be encountered.

We find that the model's main prediction of outlier blindness holds in cases of this kind as well. Suppose that an investor must choose between investments that promise two different possible returns x_1, x_2, but that the reward for choosing a particular investment is not the monetary return x but a utility $u(x)$, where $u(x)$ is an increasing but nonlinear function, perhaps with $u'(x)$ becoming particularly large for large negative values of x. The problem of efficient coding in this case is mathematically identical to the expected-value maximization problem discussed above, except that the "x" in the previous problem (the state variable, the expected value of which is to be maximized) is now the utility associated with a particular return, not the monetary value of the return. Thus, condition (2) still applies if the function m described there is understood to encode the level of utility rather than the financial return. Since the utility is itself a monotonic function of the financial return, we can again write the mean of the distribution from which the internal representation r is drawn as an increasing function $m(x)$ of the financial return. In this case, the optimal rule for encoding the financial return will satisfy the more general relation

$$m'(x) \sim u'(x)^{1-\alpha} \cdot f(x)^{\alpha}, \tag{3}$$

where, again, $\alpha = 2/3$.[14]

If the prior distribution for x has a long left tail, with larger negative values occurring with progressively smaller prior densities, the factor $f(x)^{\alpha}$ will become small for large negative outliers. But if, on the other hand $u'(x)$ also becomes larger, the more negative is x, this would be a reason for it to be efficient to discriminate *more* sharply, rather than less sharply, between larger negative values of x. These two considerations cut in opposite directions, and the marginal utility effect would tend to undercut the logic of outlier blindness. Nonetheless, the rate at which marginal utility increases will have to be quite severe for the prediction that efficiency requires outlier blindness to be overturned. Suppose, for example, that the prior $f(x)$ is a Gaussian distribution, while for negative x, $u(x)$ is proportional to $-(-x)^{\gamma}$ for some $\gamma > 1$. Then, no matter how large γ is, it will still be the case that, for large negative values of x, $m'(x)$ will decrease to zero as x becomes more negative. Thus, our model still implies outlier blindness, even if the rate at which $m'(x)$ shrinks for large negative returns is somewhat reduced by the consideration that (for ranges of values with equal prior densities) it is desirable to be able to make sharper discriminations over a range of values for which marginal utility is greater.

4.1.2. Perceptual adaptation, fast and slow

Tests of the implications of efficient coding in sensory domains often treat the distribution of sensory features in the environment as something that can be regarded as fixed. But, in financial economics applications, it is less obvious that the distributions from which random variables are drawn do not change over time; and, to the extent that they do, this raises the following question: does adaptation occur fast enough for the brain to be able to implement efficient coding even with little experience with the frequency distribution associated with a transient context?

Discussions of processes through which adaptation could occur on the basis of successive observations in a new environment are only recent, and they often imply that a very large number of new samples should be needed in order for the new prior distribution to be learned with much accuracy.[15] If that is true, then one should not expect outlier blindness to prevail in changing environments.

Our model, however, suggests the opposite, namely, that outlier blindness could prevail even if the environment is fast changing. Indeed, in our model, the decision maker can learn about the prior distribution by observing a few successive draws from it, as seems to be the case for the participants in the foregoing "five-trial adaptation" experiment. This can occur because all aspects of the encoding function $m(x)$ need not be reoptimized each time a decision maker's short-term context changes. Rather, $m(x)$ is reoptimized within a restricted family of possible functions, that vary in only one or two dimensions. Adaptation to a new context thus only requires adjustment of those one or two parameters. Given this, it can occur quite rapidly, after only a few observations drawn from the new frequency distribution.

To formalize this idea, we express the encoding function at any point in time in the form $m(x) = \tau(\phi(x))$, where

$$\phi(x) = \alpha + \beta x$$

is an affine transformation of the state space (with α an arbitrary constant and $\beta > 0$), and the "template function" $\tau(\phi)$ is a continuous, nondecreasing function with range $[\underline{m}, \bar{m}]$ that is assumed to adjust only slowly over time. Over any epoch over which the template function remains essentially constant, the above equation defines a two-parameter family of possible encoding rules (parameter-

[13] Let $L(x^{chosen}, x^{not})$ denote the loss function, where x^{chosen} is the monetary amount that is judged to be larger and x^{not} is the monetary amount judged to be smaller. The functions considered above are decreasing in the true difference $x^{chosen} - x^{not}$, but independent of the two absolute magnitudes. A second category of translation-invariant functions considered above is $L(\hat{x}, x)$, where x is the true magnitude and $\hat{x}$ the decision maker's estimate, that is increasing in the difference $|\hat{x} - x|$ but independent of the two absolute magnitudes.

[14] The same formula holds, but with $\alpha = 1/3$ if instead the objective is to minimize the mean squared error of an estimate of the utility associated with monetary return x.

[15] See, for example, Robson and Whitehead (2019) in the economics literature, or Bredenberg et al. (2020) and Aridor et al. (2020) in the literature on computational models of neural coding.

ized by α and β), all of which will be continuous, nondecreasing functions, and all of which will remain within the bounds $[\underline{m}, \bar{m}]$. The parameters α and β adjust quickly to a new short-run context, so that the expected probability of a correct choice (under the current short-run prior $f(x)$) is maximized within the two-parameter family of encoding rules, consistent with the current template function.

One implication of the model is that if the distribution of values of x shifts from $f(x)$ to $\tilde{f}(x) = f(\tilde{\phi}(x))$, where $\tilde{\phi}(x)$ is an affine transformation of the kind specified above, the optimal encoding rule for this new environment will be given by $\tilde{m}(x) = m(\tilde{\phi}(x))$. That is, only two parameters of the frequency distribution (the mean and standard deviation) need to be updated for the perceptual system to adjust to a shift in the statistics of the environment. Hence, even a small sample of observations from a new environment might well be sufficient for adaptation to occur. Our experimental results in the five-trial adaptation experiment provide behavioral evidence that this is indeed what happens. There is also neural evidence that patterns of neural activity evoked by given sensory stimuli appear to quickly adjust to the range of stimuli in a given experiment.[16]

In our proposed theory (explained further in Appendix B), the template function is also optimized in response to the decision maker's experience, but at a much lower frequency, allowing finer grained statistics of the environment to be learned. This explains why the template function should have a sigmoid shape (as assumed in Fig. 2) in the case of our application to the encoding of financial returns.

4.2. Tail risk neglect

A key practical implication of outlier blindness for financial decision-making is that it leads the decision maker to perceive extreme values as less extreme than the reality. To explain how outlier blindness leads to such "conservative bias," we need to complete our model by specifying how the investor estimates the magnitude of a given return based on its internal representation r.[17]

We assume that the brain produces an *optimal* estimate of the presented magnitude based on the information contained in r.[18] Here optimality is judged with reference to a particular loss function associated with inaccurate estimates, and a particular prior distribution over the true magnitudes that could be encountered. Because r is not perfectly informative about the true magnitude x, the resulting perceived values are not only generally incorrect, but often also biased on average.[19]

To be more specific, let $\hat{x}$ denote the investor's estimate of return magnitude x on the basis of the internal representation r. The rule for optimal estimation, in the sense of minimizing the mean squared error of the estimate, makes the estimate equal to the posterior mean for x. That is,

$$\hat{x} = E[x|r],$$

where the conditional expectation is computed using the joint distribution for x and r implied by the frequency distribution $f(x)$ from which the return is drawn and the conditional probabilities $p(r|x)$ from the efficient coding scheme described above.

In the absence of any noise in perception (i.e., if the standard deviation of $p(r|x)$ is null), $\hat{x}(r) = m^{-1}(r)$, so that $\hat{x} = x$ for all x, even extreme outliers. In such a case, there is no bias in estimated magnitudes, regardless of the nature of the transformation $m(x)$. That is, the brain should be able to correct or "undo" any perceptual bias coming from $m(x)$.[20] Instead, if encoding is noisy, as assumed in our model above, optimal estimates will not only be noisy, but will generally be biased. The bias is furthermore of a particular form (at least globally): it will be a "conservative bias," in the sense that the bias

$$b(x) = E[\hat{x}|x] - x$$

will be negatively correlated with the value of x.

This conservative bias is familiar to financial economists from the theory of linear regression: if the prior $f(x)$ is Gaussian, and $m(x)$ is a linear function, so that the joint distribution of x and r is a bivariate Gaussian distribution, then $\hat{x}(r)$ will be just the fitted value from a linear regression of x on r, and the joint distribution of x and $\hat{x}$ will also be bivariate Gaussian. The bias $b(x)$ is then a negative multiple of $x - \mu$; it grows as x is made more extreme in either direction, but only linearly in the distance from the mean.

However, the linear case is not the one of primary interest. In such a case, the degree of discriminability of nearby magnitudes (as measured by the ratio of $m'(x)$ to σ) remains constant over the entire range of possible magnitudes, and the bias $b(x)$ grows as x is made more extreme, but only linearly with $x - \mu$. Instead, with a sigmoid encoding function of the kind predicted by our model (see Fig. 2), the degree of discriminability falls to zero as x is made more extreme. This reduces the signal-to-noise ratio in the internal representation as one moves farther out into the tails of the prior distribution, and increases the extent to which the optimal estimate of x is biased toward the prior mean. As a consequence, the bias $b(x)$ grows

[16] For example, Tobler et al. (2005); Padoa-Schioppa and Assad (2008), and Soltani et al. (2012).

[17] The key point to make here is that the noisy internal representation r is *not* itself the perceived return magnitude. Rather, it is the information that the investor's brain uses to produce an estimate of return magnitude. In the neuroscience literature, this is often referred to as the problem of "decoding" the noisy internal representation.

[18] In this last respect, our model reflects discussions in Wei and Stocker (2015, 2017); Rustichini et al. (2017), and Polania et al. (2019), but differs from the discussion of the implications of adaptive coding in Bhui and Gershman (2018) and Robson and Whitehead (2019).

[19] See Woodford (2020) for further discussion of perceptual biases resulting from Bayesian decoding.

[20] Through a similar argument, Rustichini et al. (2017) propose a model of efficient coding that involves range adaptation with a linear template function (of the kind discussed in the previous subsection), but in which optimal decoding "undoes" the effects of range adaptation, so that it has few behavioral consequences. We assume optimal decoding, as they do, but show that non-trivial biases remain in our case, owing to the nonlinearity of our encoding rule.

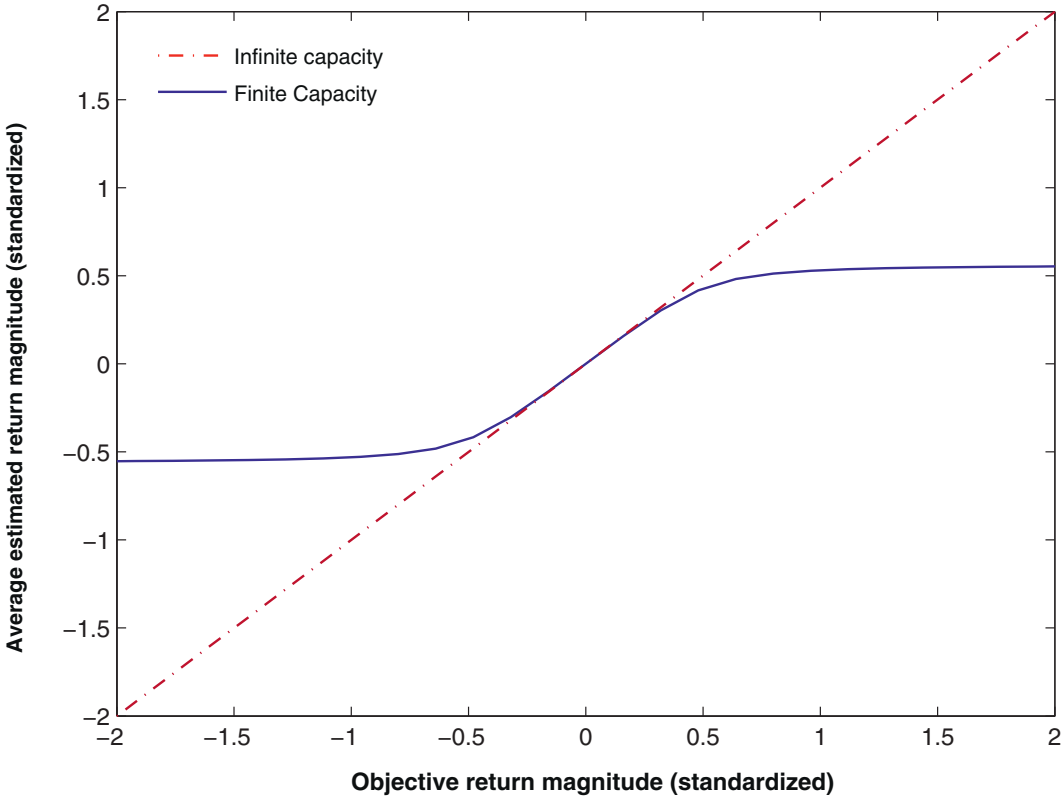

Fig. 8. Tail risk neglect. The graph shows the objective (45-degree line) vs. average estimated (sigmoid curve) return magnitude. The investor fails to appreciate the extremeness of outliers on both tails, with extremely small values being estimated as larger than actual, and extremely large values being underestimated.

more rapidly, and becomes much more important for magnitudes drawn from the tails of the prior distribution.

This is illustrated in Fig. 8.[21] The figure shows how the average estimated magnitude $E[\hat{x}|x]$ varies with x, when the prior $f(x)$ is again a Gaussian distribution $N(\mu, \omega^2)$, but the encoding function $m(x)$ is nonlinear (it is a sigmoid function with bounded range; see Fig. 2).[22] While the bias is again negatively correlated with x (it is a large negative quantity for all x more than one standard deviation greater than the prior mean, and a large positive quantity for all x more than one standard deviation less than the prior mean), it is no longer a linear function of x. The bias is negligible for values of x within one-half a standard deviation[23] of the prior mean, but is instead substantial in the case of outliers in either direction. Indeed, as Fig. 8 illustrates, $E[\hat{x}|x]$ remains within a bounded range, no matter how extreme the true magnitude x.[24]

While the numerical results shown in Fig. 8 assume that the estimate minimizes the mean squared error, it is important to note that it is not necessary that the model assumes this particular loss function to generate the conservative bias result. The more general point is that an optimal decision rule makes $\hat{x}$ a function of the posterior distribution $p(x|r)$ implied by a given representation r, and extreme outliers will be perceived as much less extreme than they actually are. The reason is simply that the person's finite representational capacity does not allow truly extreme outliers to be perceived much differently than outcomes that are somewhat unusual, but not as far out in the tail of the prior distribution to which the person's perceptions have been adapted. Therefore, if the person's estimate of the risks associated with a particular situation is based on these perceived values, then the weight given to truly extreme events will be underestimated, not because the tail events are not included in the sample on the basis of which risks are estimated, but because they are perceived to be less extreme than they actually are.

[21] The figure is adapted from Woodford (2012b). For more computational details and an extensive discussion, see Woodford (2012b) and Polania et al. (2019).

[22] The figure actually plots $E[\hat{z}|z]$ as a function of z, where $z = (x - \mu)/\omega$ is the "standardized" value of x. One can show that this normalized version of the figure is independent of the values of μ and ω, and thus applies to all Gaussian priors. The curve shown assumes a particular value of K, the bounded representational capacity of the neural system. As shown in Woodford (2012a), it has a similar sigmoid shape for other finite positive values of K.

[23] The range of values of x for which the bias is approximately zero depends on the value of K; it becomes wider as K is increased, eventually approaching the entire real line as K is made infinite (the "infinite capacity" case shown in Fig. 8).

[24] Again, these bounds depend on the value of K. They become wider as K increases, eventually exceeding any finite quantity for large enough K.

As discussed in the introduction, this provides a potential explanation for the apparent underestimation of tail risk by many investors in the period immediately prior to the global financial crisis of 2008. Large negative returns (points well to the left in Fig. 8) are predicted to be estimated on average as less negative than they are. Moreover, the size of the bias depends not on the absolute size of the returns in question, but on their standardized value: the number of standard deviations below the prior mean the contemplated return happens to be. A given size of negative return will therefore be more of an outlier (and underestimated more severely) when perceptions have adapted to a distribution of returns that is higher on average and/or less volatile. Thus, the model explains why negative returns of a given (objective) magnitude would be perceived as less damaging following a prolonged period of macroeconomic and financial stability, as proposed by Minsky (1986).

4.3. Additional factors affecting the perception of tail risk in decisions from description

An investor's mental representation of the distribution of possible outcomes to a particular investment has two elements: (1) their subjective assessment of the magnitude of each possible outcome (which is our focus above), and (2) their subjective assessment of the probability of each outcome. The nature of (2) differs depending on the kind of decision faced by the investor. As stressed in the introduction, there are two categories of decisions that investors could face: on the one hand, those in which they are not told the probabilities of the different possible outcomes and must learn them by calling to mind a sample of possible outcomes ("*decisions from experience*"); on the other hand, the so-called "*decisions from description*," in which they are explicitly told the outcome probabilities. Although the former kind of situation is our main focus in this study, it is important to also consider what should happen in the case of a decision from description, given that we agree that both cases are relevant to financial decision-making, as stressed by Barberis (2013).

A key point to make is that, for the case of decisions from description, (2) has to come from forming an internal representation of the numerical information provided about the probability of each possible outcome. It thus makes intuitive sense to formalize outlier blindness not only for the perception of outcome magnitude but also for the perception of outcome probability in this case.

We show in Appendix C.2 that applying our theory of outlier blindness to the perception of outcome probability, specifically, the joint hypothesis of efficient coding and optimal decoding of probability information, implies a conservative bias in the probability estimates of the same kind as the bias described above for the numerical information about outcome magnitudes. In the case of outcome probability, "conservative" estimates mean that small probabilities will not be perceived to be as small as they really are, while at the same time the probabilities of near-certain outcomes will not be perceived to be as close to one as they really are. This consideration provides a reason for extreme outcomes to receive excessive weight in the evaluation of a risky gamble in the case of decisions from description; it thus provides a neurobiological foundation for the kind of biases reflected in the probability weighting function posited by Prospect Theory.[25]

Outlier blindness thus has two opposing effects on investor perception in the case of decisions from description: on the one hand, it leads to underweighting extreme events, as a result of the conservative bias in the perception of outcome magnitude; on the other hand, it leads to overweighting them, as a result of the conservative bias in the perception of outcome probability. Which of these factors outweighs the other will depend on fine details of the risks being evaluated and the prior distribution from which they are assumed to have been sampled. For example, the evidence observed in Kahneman and Tversky (1979) supports the idea that, in the case of simple gambles involving a small probability of a relatively extreme outcome, the second factor dominates, leading to overweighting extreme events.

This contrasts with the case of decisions from experience, in which the assessment of outcome probability does *not* come from forming an internal representation of the numerical information provided about outcome probability; rather, it has to come from an entirely different process, namely learning (sampling outcomes). Therefore, in this case, outlier blindness only applies to the assessment of outcome magnitude, resulting in the underweighting of extreme events described above. And, indeed, laboratory evidence suggests that in the case of decisions from experience, low-probability outcomes are underweighted rather than being overweighted (e.g., Hertwig et al., 2004; Ungemach et al., 2009; Erev et al., 2010).

In this way, our theory of outlier blindness can make sense of both the pattern of underweighting rare events in decisions from experience, and that of overweighting them in decisions from description. Thus, it allows a unified understanding of tail risk neglect on the one hand, and investor overweighting of small-probability events on the other, two important ideas from prior work that at first glance could seem contradictory.

Another possible reason for overweighting of small-probability "outlier" outcomes in the case of decisions from description could be that greater attentional focus is directed to these particular outcomes, as postulated by the theories of Bordalo et al. (2012) and Koszegi and Szeidl (2013). Our theory can be extended to allow for effects of differential attentional focus on particular contingencies, if we suppose that the capacity K allocated to the encoding of returns is greater for the returns associated with certain contingencies. If the contingencies under which extreme returns occur are the ones to which more cognitive resources are devoted, this could be a further reason for extreme outcomes to be over-weighted in decisions from description.

By contrast, there is much less reason to expect salience effects of this kind in the case of decisions from experience in which the investor simply happens to have observed dif-

[25] It should be noted that, in our theory, the degree of probability weighting should vary depending on the prior distribution to which preferences have been adapted.

ferent returns on different occasions without having had reason to treat some outcomes as instances of a different category. Hence, one would not expect salience effects to outweigh the effects of outlier blindness in the case of decisions from experience.

5. Discussion

To summarize, we propose a novel model of efficient coding that, relative to prior efficient coding models, is both simpler and more general in dimensions that are highly relevant for financial economics. It is simpler in that the model abstracts from any specific biophysical model of how neurons represent environmental features. At the same time, it is more general in that it considers a wide range of possible criteria for the efficiency of a coding scheme. In our benchmark case, the one that corresponds to the incentives provided in our experiment, efficiency is defined as maximizing the accuracy with which two return values are discriminated. But, we show that the core prediction of our model also obtains when the reward for a correct decision varies across different cases and when the decision is more complex than a simple choice between two options. It also holds whether or not the investment context is assumed to be stable, and even if it is assumed to be affected by frequent regime shifts (as is sometimes the case in financial trading, for example).

The collection of experimental findings strongly supports the key prediction of our model, which is the idea that investors can make fine distinctions between values in the expected range but only coarse distinctions (if any at all) between outliers. Our model also suggests that investor misperception of extreme values could be particularly pronounced in the case of fat-tailed returns.

That said, it is worth emphasizing certain limitations of our study. One important limitation is that our analysis does not enable us to say how long a sudden new development should continue to be surprising—and hence misperceived because it is out of line with the distribution of values to which people's perceptions are attuned—if the situation persists. How long does the initial misperception caused by outlier blindness last? We think it is plausible that, under some circumstances, macroeconomic developments can be misperceived for months because they are outliers relative to people's expectations. Likewise, we think it is plausible that traders can underreact to market shocks (e.g., large volatility movements) for several days due to outlier blindness. Yet, our theory by itself does not answer this question. It must be augmented by a theory of learning and, above all, further empirical study of the determinants of the speed with which the reallocation of limited perceptual capacities occurs in different contexts.[26]

It would be interesting to study in the future whether outlier blindness causes distortions of asset prices, and to quantify its portfolio implications for practitioners. For one example, the current study, along with recent evidence of initial underreaction to volatility shocks (Lochstoer and Muir, 2021), suggests that traders could initially underestimate the size of unexpected volatility movements, possibly leading them to stay on the wrong side of the trade when such moves happen. Data on the VIX and tradable futures could be used to examine this possibility and also measure the size of trader underreaction, the ensuing portfolio losses, and how long it takes the traders to catch up.

Finally, our analysis does not preclude the possibility that financial agents will consciously correct for outlier blindness in their decision-making subsequent to dissemination of this research, just as the dissemination of research on stock market "anomalies" has reduced the extent to which they are observed (McLean and Pontiff, 2016). Some have indeed pointed out that, while it is not possible to prevent our visual system from being subject to visual illusions, learning about these illusions makes it possible for people to consciously correct their decisions.[27] We do not mean that an intellectual understanding of outlier blindness should prevent outliers from being perceived as less extreme than they are; efficient coding should be understood as being applied subconsciously and optimized through evolution, rather than through conscious reasoning. Nonetheless, we believe that a better understanding of characteristic misperceptions of the magnitude of financial returns can contribute to investors learning how to consciously improve their decisions, by realizing that extreme outcomes could well be more severe than one's intuitive assessment of them would indicate.[28]

Appendix A

Here we explain our model in more detail. We begin with a discussion of the consequences of eventual fully optimal adaptation to a given prior distribution, without asking how long it might take for that to occur. We first show why outlier blindness should occur under the assumptions of our baseline model in Section 1, and then derive the consequences of the generalizations discussed in Section 3.1. We then discuss our extended model with two different speeds of adaptation, in which range adaptation occurs rapidly but adaptation of the "template function" takes longer.

Optimal adaptation to a single prior

We first discuss the implications of efficient coding under the assumption that the magnitudes to be encoded (shades of grey, in our experiment, or financial returns, in the application to an investment problem) are always drawn from a single prior distribution, and that the coding

[26] Note that our theory is not alone in leaving this crucial question open. Any theory according to which people's degree of sensitivity to some state depends on how it compares with expectations raises a similar question. The theory of reference-dependent preferences of Koszegi and Rabin (2006) and the diagnostic expectations of Gennaioli and Shleifer (2018) are two other examples.

[27] For example, Ariely (2010).

[28] We do not view the increasing automation of trading as a "remedy" for the case of trading, insofar as automation only concerns the mechanical aspects of trading; humans continue to control tasks that involve judgment (such as key trading decisions, forecasting future prices, and gauging whether price levels are correct).

scheme is optimized for this prior. Let each of two magnitudes x_i (for $i = 1, 2$) be independent draws from a prior distribution with probability density function $f(x)$ and associated cumulative distribution function $F(x)$. The support of this distribution is assumed to be the real line (the magnitudes are scalar-valued). We further assume that the prior distribution is atomless, so that $F(x)$ is a continuous nondecreasing function.

We consider imprecise encoding schemes in which each magnitude x_i has a noisy internal representation r_i. Conditional on the magnitudes (x_1, x_2), the distribution of each of the r_i is independent of the other, and is given by a Gaussian distribution

$$r_i \sim N(m(x_i), \sigma^2),$$

the mean of which depends on the true magnitude x_i, and the variance of which is the same for all possible magnitudes. The encoding function $m(x)$ is assumed to be a nondecreasing function,[29] defined on the entire support of the prior distribution, and taking values in a bounded interval $[\underline{m}, \bar{m}]$. As explained in the main text, it is this bound on the range of the function $m(x)$ (together with the positive value of σ) that implies a necessary degree of imprecision in internal representations.[30]

A1. Optimal encoding under our baseline objective

In our baseline model, the decision maker's task (as in our experiment) is to determine which of the two magnitudes is larger; we assume that this decision must be based on the pair of internal representations (r_1, r_2). Our theory of efficient coding proposes that both the encoding function $m(x)$ and the decision criterion (giving a response as a function of (r_1, r_2)) are chosen so as to maximize the probability of the decision maker's making a correct judgment (that is, a judgment that item 1 is greater, when $x_1 > x_2$, or that item 2 is greater, when $x_2 > x_1$). (Equivalently, we can assume that the probability of an incorrect judgment is minimized.) Here the probability of a given outcome is computed under the assumption that x_1, x_2 are independent draws from the prior distribution, and the optimization is conditional upon the given bounds $[\underline{m}, \bar{m}]$ and the given standard deviation σ for the encoding noise.

Under our assumptions, it can be shown that, in the case of any nondecreasing function $m(x)$, the response "magnitude 1 is greater" is more likely to be the correct response if and only if $r_1 > r_2$.[31] Hence, we can assume that this is the response rule (as already proposed in the main text), and consider the optimal choice of the function $m(x)$.

For any encoding function $m(x)$, it then follows that, for any true magnitudes (x_1, x_2) with $x_2 > x_1$, the probability of an incorrect response (i.e., the response "magnitude 1 is greater") will equal

$$\text{Prob}[r_1 > r_2 \,|x_1, x_2] \;=\; \Phi\left(\frac{m(x_1) - m(x_2)}{\sqrt{2}\sigma}\right), \tag{A.1}$$

where $\Phi(z)$ is the cumulative distribution function of a standard normal distribution.[32] More generally, for any (x_1, x_2), the probability of an incorrect response will equal

$$\Phi\left(-\frac{|m(x_1) - m(x_2)|}{\sqrt{2}\sigma}\right).$$

The overall probability P^e of an erroneous judgment is then given by

$$P^e \;=\; \int\int f(x_1) f(x_2)\, \Phi\left(-\frac{|m(x_1) - m(x_2)|}{\sqrt{2}\sigma}\right) dx_1 dx_2. \tag{A.2}$$

The efficient coding hypothesis requires that the function $m(x)$ be chosen (among all nondecreasing functions satisfying the specified bounds) so as to minimize P^e for a given prior distribution $f(x)$.

Under our assumption that $F(x)$ is continuously increasing, we can alternatively parameterize a magnitude x_i by its quantile $q_i = F(x_i)$ in the prior distribution, and equivalently specify its noisy internal representation as

$$r_i \;=\; \underline{m} + (\bar{m} - \underline{m}) \cdot \tilde{r}_i,$$

where $\tilde{r}_i$ is an independent draw from a distribution $N(n(q_i), \nu^2)$, and we define

$$n(q) \;\equiv\; \frac{m(F^{-1}(q)) - \underline{m}}{\bar{m} - \underline{m}}, \quad \nu \;\equiv\; \frac{\sigma}{\bar{m} - \underline{m}}. \tag{A.3}$$

Note that, under this alternative measure of the magnitude, q_i necessarily lies in the interval $[0, 1]$, and is uniformly distributed over that interval under the prior, and $n(q)$ is a nondecreasing function defined on the interval $[0, 1]$ that also has a range equal to $[0, 1]$. Thus, any encoding function $m(x)$ with range $[\underline{m}, \bar{m}]$ corresponds to a normalized encoding function $n(q)$ with range $[0, 1]$. In terms of the normalized encoding function, the probability of an error can be expressed as

$$P^e \;=\; \int\int \Phi\left(-\frac{|n(q_1) - n(q_2)|}{\sqrt{2}\nu}\right) dq_1 dq_2. \tag{A.4}$$

Note that expression (A.4) involves no parameters of the model other than ν (which is just the reciprocal of

[29] In fact, there is no loss of generality involved in restricting our attention to nondecreasing encoding functions, since the use of a nonmonotonic function would not allow one to achieve a lower value of the expected loss criterion discussed below. Because this result is intuitive, we simplify the discussion here by directly assuming a nondecreasing function.

[30] This way of modeling the constraints on the precision of internal representations is also used by Steiner and Stewart (2016); Polania et al. (2019), Robson and Whitehead (2019), and Zhang et al. (2020), among others. The quantity K, proposed in the main text as a measure of the capacity for differentiation of alternative magnitudes, is called the "Thurstone invariant" by Zhang et al. (2020), who present evidence that it is a subject-specific magnitude with a similar value for different cognitive tasks.

[31] Here we neglect the case of an exact tie in which $r_1 = r_2$, as this will occur with probability zero, regardless of the encoding function chosen.

[32] Note that since r_1 and r_2 are both normally distributed (conditional on the x_i), their difference $r_1 - r_2$ is normally distributed as well, with mean $m(x_1) - m(x_2)$ and variance $2\sigma^2$.

the ratio $K \equiv [\bar{m} - \underline{m}]/\sigma$ defined in the main text). Furthermore, the set of admissible functions $n(q)$ is also independent of the model parameters. Thus, the feasible degree of precision of comparative judgments depends only on the ratio K (or alternatively the normalized noise parameter ν), and not on the values of the parameters $\underline{m}, \bar{m}$, or σ individually, as stated in the main text.

It is even more noteworthy that the objective (A.4) can be expressed in a way that is independent of the prior distribution. This immediately gives us the following key result.

Proposition 1. For a given value of the parameter $\nu > 0$, let $n^*(q)$ be the nondecreasing function mapping the interval $[0, 1]$ to itself that minimizes the value of P^e defined in (A.4). (Note that the solution to this problem is independent of the prior $F(x)$.) Then, for any prior $F(x)$, the encoding rule $m(x)$ required for efficient coding is given by

$$m(x) \;=\; \underline{m} + (\bar{m} - \underline{m}) \cdot n^*(F(x)), \tag{A.5}$$

Proof. It follows directly from the fact that $n^*(q)$ minimizes expression (A.4) that the corresponding encoding function $m(x)$ given in (A.5) will minimize (A.2). Thus, this function $m(x)$ solves our efficient coding problem. □

A2. Context-sensitivity of the efficient encoding rule

Proposition 1 shows that the efficient encoding rule $m(x)$ will depend on the prior distribution $F(x)$, as discussed in the main text. In particular, it implies that one of the consequences of efficient coding is range normalization.

Corollary 1. Let $F(x)$ and $\tilde{F}(x)$ be the CDFs corresponding to two possible prior distributions for the magnitude x, such that there exists an increasing affine transformation of the state space,

$$\phi(x) \;\equiv\; \alpha + \beta x \tag{A.6}$$

with $\beta > 0$, with the property that

$$\tilde{F}(x) \;=\; F(\phi(x))$$

for all x. And, suppose that $m(x)$ and $\tilde{m}(x)$ are the optimal encoding functions in the case of priors $F(x)$ and $\tilde{F}(x)$, respectively. Then,

$$\tilde{m}(x) \;=\; m(\phi(x)) \tag{A.7}$$

for all x. That is, $\tilde{m}(x)$ can be obtained from $m(x)$ by shifting and stretching the state space in the same way as is required to obtain $\tilde{F}(x)$ from $F(x)$.

This result generalizes the familiar idea of range normalization (more below). The conventional concept assumes a stimulus distribution with a bounded range, and assumes that a second distribution is obtained from an original distribution by stretching the range. Range normalization asserts that, when the range of the distribution is stretched, the encoding function should be correspondingly stretched, so that the new range of stimuli are mapped into the same range of internal representations. But, as our result shows, the idea can be extended to cases in which the support of the distribution need not be bounded. We can still assert that if a second distribution is obtained from some first distribution through an affine transformation of the state space, the appropriate encoding function in the case of the second distribution of magnitudes should be obtained from the appropriate function in the case of the first distribution through that same affine transformation.

As a simple example, our Corollary implies that if a second prior distribution is a simple horizontal translation of some first distribution, as in the case of the prior distributions $f(x)$ and $\tilde{f}(x)$ shown in the upper right panel of Fig. 2, then the optimal encoding function in the case of the second prior will also be a corresponding horizontal translation of the optimal encoding function for the first prior, as shown in the lower right panel of the figure.

A further step in the reasoning is required to explain the prediction illustrated in Fig. 2, which is in turn the key prediction tested (and confirmed) in our experiment. This is the assumption in the figure that a unimodal prior distribution $f(x)$ is associated with an optimal encoding function $m(x)$ with a sigmoid shape, with the steepest part of the sigmoid curve around the peak of the function $f(x)$. We consider this issue next.

Let x_q be the magnitude with quantile q in the prior distribution, i.e., such that $F(x_q) = q$. And, suppose that $n^*(q)$ is differentiable at this value of q.[33] It then follows directly from (A.5) that the slope of the optimal encoding function at this point will be given by

$$m'(x_q) \;=\; (\bar{m} - \underline{m}) \cdot n^{*'}(q) \cdot f(x_q). \tag{A.8}$$

It is this slope that determines the discriminability of different magnitudes near x, as discussed in the main text (and illustrated in Fig. 2). If we compare the optimal encoding rules for two different prior distributions, and in each case consider the slope $m'(x_q)$ at the value corresponding to quantile q, we find that $m'(x_q)$ will be smaller for the distribution that has the smaller prior density $f(x_q)$ at that quantile. Furthermore, the slope will be smaller exactly in proportion to the extent that the prior density is smaller.

Now, let q be a quantile that is relatively extreme (so that x_q is an outlier). The greater the extent to which the prior distribution has a long tail, so that the density $f(x_q)$ is small for all sufficiently extreme values of q, the clearer it is that the optimal encoding rule must be relatively flat at those points in the prior distribution.

A3. A small-noise approximation

But, we can go farther and show that, except when the capacity K is quite small, optimal encoding requires that the slope $n^{*\prime}(q)$ not vary much across different quantiles. This result can be stated in a particularly strong (and simple) form in the case in which the encoding noise parameter ν is small (that is, in the limiting case in which relatively precise discrimination is possible).

Suppose that we fix a function $n(q)$, which we shall assume is monotonically increasing and continuously differ-

[33] This will be true of most q, though it could fail to be true on a set of measure zero.

entiable, but vary the size of ν. Let us first consider the value of the inner integral in (A.4) as a function of q_2. For small enough values of ν, the value of this integral can be approximated as follows:

$$\begin{aligned}
&\int \Phi\left(-\frac{|n(q_1)-n(q_2)|}{\sqrt{2}\nu}\right) dq_1 \\
&= \int q'(n_1)\,\Phi\left(-\frac{|n_1-n_2|}{\sqrt{2}\nu}\right) dn_1 \\
&= \sqrt{2}\nu \cdot \int_{-n_2/\sqrt{2}\nu}^{0} q'(n_2+\sqrt{2}\nu\xi)\,\Phi(\xi)\,d\xi \\
&\quad + \sqrt{2}\nu \cdot \int_{0}^{(1-n_2)/\sqrt{2}\nu} q'(n_2+\sqrt{2}\nu\xi)\,\Phi(-\xi)\,d\xi \\
&\approx \sqrt{2}\nu\, q'(n_2) \cdot \left\{ \int_{-n_2/\sqrt{2}\nu}^{0} \Phi(\xi)\,d\xi + \int_{0}^{(1-n_2)/\sqrt{2}\nu} \Phi(-\xi)\,d\xi \right\} \\
&\approx 2\sqrt{2}\nu\, q'(n_2) \int_0^{\infty} \Phi(-\xi)\,d\xi \;=\; \frac{2\nu}{\sqrt{\pi}} \frac{1}{n'(q_2)}
\end{aligned} \tag{A.9}$$

Here the first line follows from a change of the variable of integration, defining the function $q(n)$ as the inverse of the function $n(q)$ and letting $n(q_1) = n_1, n(q_2) = n_2$; and the second line follows from another change of variables, in which we define $\xi \equiv (n_1 - n_2)/\sqrt{2}\nu$ and break the integral into two parts, corresponding to the domains over which ξ is negative or positive. The third line then follows from a local approximation of the function $q'(n)$, in which we omit terms smaller than order ν, and the fourth line from another approximation (extending the range of integration to infinity) that also contributes only terms that are smaller than order ν. The resulting expression (A.9) is therefore accurate to order ν for any value of q_2.

Integrating expression (A.9) over q_2, we then obtain an approximation to the objective (A.4),

$$P^e \;\approx\; \frac{2\nu}{\sqrt{\pi}} \int_0^1 \frac{dq}{n'(q)}$$

again accurate to order ν. Thus, the normalized encoding function $n(q)$ that will achieve the lowest possible error probability, at least in the case of small enough values of the noise parameter ν, will be the function with the lowest value of the integral

$$\int_0^1 \frac{dq}{n'(q)}. \tag{A.10}$$

Because of the bounded range of the function $n(q)$, its derivative $n'(q)$ must satisfy the integral equation

$$\int_0^1 n'(q)\,dq \;=\; 1. \tag{A.11}$$

Thus, the optimal slope function $n'(q)$ (in the small-noise approximation) will be the one that maximizes (A.10) subject to the constraint (A.11). It is easily verified that the solution to this problem is given by $n'(q) = 1$ for all q. Given that $n(q)$ is restricted to the range $[0, 1]$, we see that the optimal encoding function is given by

$$n^*(q) \;=\; q \tag{A.12}$$

for all $0 \le q \le 1$. Combining this result with (A.5) then yields the following.

Proposition 2. Let the prior $F(x)$ be fixed, and consider the implications of efficient coding in the case of each sequence of progressively smaller values for ν (measuring the degree of encoding noise). The sequence of optimal encoding functions $m(x)$ for different values of ν will satisfy

$$\lim_{\nu\to 0} m(x) \;=\; \underline{m} + (\bar{m} - \underline{m}) \cdot F(x). \tag{A.13}$$

That is (up to an affine transformation that does not affect the information content of the neural representation), the optimal encoding function is given by the cumulative distribution function for the prior to which the encoding is adapted.

This result (the optimality of encoding magnitudes by an imprecise record of their quantile in the prior distribution) explains the formula (1) given in the main text. The asymptotic formula (A.13) is what we assume in Fig. 2: in the figure, the prior density functions $f(x)$ and $\tilde{f}(x)$ are two Gaussian distributions (with the same variance but different means), while the associated optimal encoding rules are both affine transformations of the CDFs of these Gaussian distributions. And, it immediately gives rise to the prediction of outlier blindness. Eq. (A.13) implies that $m'(x)$ will be proportional to $f(x)$, the prior density at point x in the state space. If the prior distribution is unimodal, with $f(x)$ monotonically decreasing as x moves farther away from the prior mode in either direction, it follows that $m'(x)$, which determines the accuracy with which nearby magnitudes can be distinguished at that part of the state space, should also be monotonically decreasing the farther x is from the prior mode, i.e., the greater the extent to which x is an outlier.

A4. Comparison with the literature

A number of papers in the neuroscience literature discuss an implication of efficient coding that is closely related to Eq. (1), though not expressed in terms of the derivative of a function $m(x)$ that maps stimulus magnitudes to the so-called "Thurstone scale."[34] It is common to consider the implications of efficient coding for the acuity of perception as measured by the "discrimination threshold," namely, the size of change in the stimulus magnitude required for a second stimulus to be recognized to be larger than some reference stimulus with a certain probability. In our model, the probability that a magnitude x will be judged greater than some reference magnitude x_0 is equal to

$$\Phi\left(\frac{m(x) - m(x_0)}{\sqrt{2}\sigma}\right),$$

using (A.1). Thus, the discrimination threshold is the size of increase in x required to increase $m(x)$ by some fixed amount; it varies inversely with the slope of the encoding function $m(x)$. Hence, Eq. (1) implies that measured discrimination thresholds should vary inversely with the prior density in a given region of the stimulus space.

This inverse relation between the discrimination threshold and the frequency distribution of stimulus

[34] While this device has long been used in the psychophysics literature as a way of modeling variation in the degree of discriminability of nearby stimuli, it is not commonly used (as we propose here) as a way of modeling the resource constraint in models of efficient coding.

magnitudes in the natural environment has been shown to be an implication of efficient coding in a number of biologically motivated models of efficient coding (e.g., Laughlin, 1981; Wei and Stocker, 2015; Ganguli and Simoncelli, 2016). These papers, however, obtain the result under different assumptions from ours. In the case of Laughlin (1981), the internal representation r is a single real number, as in our model, but it is assumed to be a deterministic function of x, rather than random; the limited precision of the representation is instead imposed by assuming that only a finite number of different values of r can be read out. The models of Wei and Stocker (2015) and Ganguli and Simoncelli (2016) instead assume a high-dimensional internal representation, indicating the degree of activation of different members of a large number of neurons in a region of the brain involved in processing the sensory feature in question. However, the information that allows different stimulus magnitudes to be discriminated can largely be summarized by variation in one dimension, as in our model.

Besides modeling the constraint on feasible internal representations differently from what we do here, neuroscience models typically assume that encoding is optimized for a different objective, namely, maximization of the Shannon mutual information between the objective magnitude x and its internal representation r. This is a property of the joint distribution of the random variables x and r that measures how informative r is about the value of x (and vice versa). This objective allows an efficient coding problem to be defined without reference to any intended use of the internal representation in making some further decision, which is convenient in applications to early stages of sensory processing, where similar initial processing is often assumed to take place regardless of the eventual higher level decisions that will be made on the basis of the sensory information. We instead assume that encoding is optimized to maximize a decision maker's success at a particular task, under the further assumption that optimal use is made of the information contained in the internal representation. (In the analysis above, the measure of success that is maximized is the probability of a correct comparative judgment, which corresponds to maximizing expected reward in our experiment. In the next section, we discuss a variety of alternative criteria that might be maximized instead, corresponding to alternative types of decision problems.)

Moreover, while the neuroscience papers conclude, like us, that the discriminability of nearby states should vary monotonically with the prior probability density, the outlier blindness result that we stress has not been a focus of that literature. One reason is that the probability distributions of sensory features that are studied in the neuroscience literature often do not have the unimodal form assumed in Fig. 2; hence, the regions of the stimulus space where discrimination is poorer are not necessarily outliers. For example, one much-studied case is the accuracy of discrimination between different orientations of lines or edges in the visual environment. It is well-established that more oblique orientations are less precisely discriminated than orientations that are nearly vertical or nearly horizontal (orientations near the so-called "cardinal" orientations). This is consistent with efficient coding, given that near-vertical and near-horizontal orientations occur much more frequently than oblique orientations, in both natural and man-made environments (Girshick et al., 2011; Wei and Stocker, 2015; Ganguli and Simoncelli, 2016). But, while oblique orientations occur less frequently, they cannot be considered more "extreme" than near-cardinal orientations. The kind of unimodal distribution that we assume in Fig. 2 is instead characteristic of financial returns (the application of particular interest to us here), as well as being the kind of distribution used in our experiment.

And, finally, the neuroscience papers have mainly focused on differences in the discriminability of different types of stimuli that are thought to be relatively stable (such as the poorer discrimination of oblique orientations, just mentioned), and to explain these in terms of relatively invariant features of the sensory environment. In Fig. 2 we have instead emphasized the way in which the theory implies that the degree of discriminability of two objective magnitudes can vary depending on the context in which they are encountered, and this is also the implication that our experiment is designed to test. In this respect as well, our emphases are different from those of much of the neuroscience literature[35] because of our interest in drawing out potential implications of the theory for financial decision-making.

A5. Efficient coding with alternative objectives

Above, we have assumed that the encoding function is optimized to maximize the probability of a correct choice, which corresponds to the incentive structure in our experiment. But, as discussed in the main text, other objectives are of particular relevance for financial applications.

Maximizing the expected value of the chosen item Suppose instead that encoding is optimized to maximize the average value (i.e., the mean value of x) of the item that is chosen. If the true magnitudes of the items offered are x_1 and x_2, the probability of an incorrect choice is again given by $\Phi(-|m(x_1) - m(x_2)|/\sigma)$, but now the loss from an incorrect choice is given by $|x_1 - x_2|$, rather than by a positive constant. (An incorrect choice is more costly when the difference in the two amounts is greater.) The expected loss in the case of an encoding function $m(x)$ is then equal to

$$\mathrm{E[Loss]} = \int\int f(x_1) f(x_2)\, |x_1 - x_2|\, \Phi \left(-\frac{|m(x_1) - m(x_2)|}{\sqrt{2}\sigma} \right) dx_1 dx_2, \tag{A.14}$$

instead of (A.2). In this case, efficient coding requires that the function $m(x)$ be chosen (among all nondecreasing functions satisfying the specified bounds) so as to minimize (A.14) for a given prior distribution $f(x)$.

[35] Exceptions include Wei and Stocker (2017), which contains discussions of the way in which the theory of efficient coding might also explain relatively high-frequency changes in the precision of discrimination when the statistics of the sensory environment change.

Just as above, we can use a small-noise approximation to obtain an explicit solution for the optimal encoding function for a given prior. When σ is small, we obtain

$$\begin{aligned}
&\int f(x_1)\,|x_1 - x_2|\,\Phi\left(-\frac{|m(x_1)-m(x_2)|}{\sqrt{2}\sigma}\right) dx_1 \\
&= \int_{\xi_{min}}^{0} \frac{\sqrt{2}\sigma}{m'(x_1(\xi))}\, f(x_1(\xi))(x_2 - x_1(\xi))\,\Phi(\xi)\,d\xi \\
&\quad + \int_{0}^{\xi_{max}} \frac{\sqrt{2}\sigma}{m'(x_1(\xi))}\, f(x_1(\xi))(x_1(\xi) - x_2)\,\Phi(-\xi)\,d\xi \\
&\approx \frac{2\sigma^2 f(x_2)}{(m'(x_2))^2}\left\{\int_{-\infty}^{0}(-\xi)\Phi(\xi)d\xi + \int_{0}^{\infty}\xi\,\Phi(-\xi)d\xi\right\} \\
&= \frac{\sigma^2 f(x_2)}{(m'(x_2))^2}. \qquad \text{(A.15)}
\end{aligned}$$

Here the first line uses the change of variable

$$x_1(\xi) \;\equiv\; m^{-1}(m(x_2) + \sqrt{2}\sigma\xi)$$

for values of ξ varying between $\xi_{\min} \equiv -(m(x_2) - \underline{m})/\sqrt{2}\sigma$ and $\xi_{\max} \equiv (\bar{m} - m(x_2))/\sqrt{2}\sigma$; the second line is the small-noise approximation, retaining only terms of order σ^2 or larger; and the final line evaluates the definite integrals.

Integrating (A.15) over the possible values of x_2 (weighted by their probability of occurrence under the prior), we find that

$$\mathrm{E[Loss]} \;\approx\; \sigma^2 \int \frac{(f(x_2))^2}{(m'(x_2))^2}\,dx_2, \qquad \text{(A.16)}$$

if we retain only terms of order σ^2 or larger. The efficient coding problem is then to choose a function $m'(x)$ to minimize (A.16), subject to the constraint that

$$\int m'(x)\,dx \;\le\; \bar{m} - \underline{m}, \qquad \text{(A.17)}$$

as required by the bounded range of the encoding function.

It is easily seen that the constraint (A.17) must bind, and that the first-order condition requires that

$$m'(x) \;\sim\; f(x)^{2/3} \qquad \text{(A.18)}$$

for all x in the support of $f(x)$, which is just the case $\alpha = 2/3$ of condition (2) in the main text. The constant of proportionality, which corresponds to the Lagrange multiplier associated with constraint (A.17), must be chosen so as to ensure that (A.17) holds with equality. This determines the function $m'(x)$. Integration of this function then determines the encoding function up to a constant of integration, which can be chosen so as to make the range of the encoding function precisely the interval $[\underline{m}, \bar{m}]$. Thus, we obtain

$$\lim_{\sigma\to 0} m(x) \;=\; \underline{m} + (\bar{m} - \underline{m}) \cdot \frac{\int_{-\infty}^{x} f(\tilde{x})^{2/3} d\tilde{x}}{\int_{-\infty}^{\infty} f(\tilde{x})^{2/3} d\tilde{x}} \qquad \text{(A.19)}$$

as an alternative to (A.13). As explained in the main text, we find that efficient coding implies outlier blindness in this case as well since the slope of the optimal encoding function is again smaller wherever the prior density is smaller.

Maximizing expected utility of the chosen item. We can generalize this result, to consider the case in which the encoding rule maximizes not the expected financial return on the investment that is chosen, but the expected utility from that financial return, where utility is assumed to be some smoothly increasing function $u(x)$ of the financial return x. Let us again suppose that the available financial returns are drawn from a prior distribution $f(x)$. This implies a prior distribution $g(u)$ for the utility that could be obtained by choosing different investments, where

$$g(u(x)) \;=\; f(x)/u'(x). \qquad \text{(A.20)}$$

Furthermore, any encoding rule $m(x)$ can be expressed as $m(x) = k(u(x))$ for some increasing function $k(u)$ that satisfies the bounds $\underline{m} \le k(u) \le \bar{m}$ for all values of u. We can then define the efficient coding problem as the choice of a function $k(u)$ to maximize the expected value of u from the chosen investment, subject to the bounds on the range of the function.

This problem has exactly the same mathematical form as the one just considered, if we replace the variable x by u, the prior $f(x)$ by $g(u)$, and the encoding function $m(x)$ by $k(u)$. The same argument as the one used to derive (A.18) implies in this case that the optimal encoding rule must satisfy

$$k'(u) \;\sim\; g(u)^{2/3}.$$

Then, noting that $m'(x) = k'(u(x))u'(x)$ and using (A.20), we see that, in this case, efficient coding requires that

$$m'(x) \;\sim\; [f(x)/u'(x)]^{2/3} \cdot u'(x) \;=\; u'(x)^{1/3} \cdot f(x)^{2/3}. \qquad \text{(A.21)}$$

This is just the case $\alpha = 2/3$ of condition (3) in the main text. It remains the case that, given any value for the marginal utility $u'(x)$, the optimal degree of sensitivity to alternative financial returns near x will be lower the lower is the prior density $f(x)$ at that value. And, as discussed in the main text, it can easily be the case that the optimal value of $m'(x)$ decreases for more extreme values of x, even if $u'(x)$ grows sharply for more extreme values of x.

Minimizing estimation error. Another case of interest is one in which an investor must make a decision on the basis of an estimate of the value of some state variable x, with losses (from a less accurate decision) that will be proportional to the squared error of the investor's estimate $\hat{x}$. In this case, the efficient coding problem is to choose an encoding function $m(x)$ and an estimation rule $\hat{x}(r)$ so as to minimize the value of

$$\mathrm{E[Loss]} \;=\; \int\int f(x)\,\frac{1}{\sigma}\phi\left(\frac{r - m(x)}{\sigma}\right)(\hat{x}(r) - x)^2\,drdx, \qquad \text{(A.22)}$$

where $\phi(z) \equiv \Phi'(z)$ is the probability density function of the standard normal distribution.

Again we can obtain an explicit solution using a small-noise approximation. For any encoding function $m(x)$, the optimal estimation rule $\hat{x}(r)$ is implicitly defined by

$$\int f(x)\,\phi\left(\frac{r - m(x)}{\sigma}\right)(\hat{x}(r) - x)\,dx \;=\; 0$$

for each value of r. (This means that $\hat{x}(r)$ should be the mean of the posterior distribution for x, given the prior

$f(x)$ and the noisy internal representation r.) In the small-noise limit, the solution to this equation can be approximated by

$$\hat{x}(r) \approx m^{-1}(r),$$

where we keep only terms of order σ or larger. Then, if we let $\xi \equiv (r - m(x))/\sigma$, and express the loss $(\hat{x}(r) - x)^2$ as a function of x and ξ rather than x and r, we obtain

$$(\hat{x} - x)^2 \approx \frac{\xi^2}{(m'(x))^2}\,\sigma^2,$$

keeping only terms of order σ^2 or larger.

We thus obtain a small-noise approximation to the loss function (A.22), given by

$$\begin{aligned} \mathrm{E}[\mathrm{Loss}] &\approx \sigma^2 \cdot \int\int f(x)\,\phi(\xi)\,\frac{\xi^2}{(m'(x))^2}\,d\xi\,dx \\ &= \sigma^2 \cdot \int \frac{f(x)}{(m'(x))^2}\,dx. \end{aligned} \tag{A.23}$$

The efficient coding problem is then to choose a function $m'(x)$ to minimize (A.23), subject to the constraint (A.17). The first-order condition for this problem is easily seen to require that

$$m'(x) \sim f(x)^{1/3}, \tag{A.24}$$

which is the case $\alpha = 1/3$ of condition (2) in the main text. Once again, we find that efficient coding implies outlier blindness, since the slope of the optimal encoding function is again smaller wherever the prior density is smaller.

Alternatively, suppose that x represents a financial return in some state of the world, and that losses depend not on the squared error of the investor's estimate of this return, but on the squared error of the investor's estimate of the utility implied by this return, where utility is a nonlinear function $u(x)$ of the return. Using the same kind of argument as is used above to derive (A.21), we can show that, in the case of a nonlinear utility function, (A.24) takes the more general form

$$m'(x) \sim u'(x)^{2/3} \cdot f(x)^{1/3}.$$

This is just the case $\alpha = 1/3$ of condition (3) in the main text. Once again, we see that it continues to be optimal to have less precise discrimination between values of x in a region of the support of the prior distribution where the prior density is smaller, so that our prediction of outlier blindness remains valid (with modest qualifications) in this case as well.

Appendix B. Fast and slow adaptation to changing environments

A key implication of our results above is that efficient coding (in the sense of optimal adaptation to some single prior) implies range adaptation. We have already noted this in the case of our baseline model (the Corollary above); in fact, a similar corollary can be established in the case of each of the other two objectives discussed above, in which the objective for which perceptions are optimized depends only on the true magnitude x (rather than on some nonlinear utility derived from x).[36] Here we discuss our proposal (in Section 3.2 of the main text) that this is not just one of the many ways in which the encoding of returns should adjust to the frequency distribution that investors encounter, but one that should occur more rapidly than other kinds of adaptation.

In our two-speed model of adaptation to environmental frequencies, range adaptation occurs more rapidly than other aspects of the adjustment of the encoding rule required by the theory of efficient coding. Here we explain this idea in the context of our baseline model (i.e., under the assumption that the objective is maximization of the probability of a correct binary comparison). This is purely for convenience; a similar two-speed model of adaptation can be formulated in either of the other cases as well. The key idea of the more general theory is the observation that the prediction of range adaptation, which is a crucial step in the argument for the prediction that we experimentally test, can be obtained under a much weaker (and more plausible) assumption than fully optimal adaptation of the encoding rule to each relatively transitory environment that is encountered.

B1. Fast range adaptation

Consider a parametric family of possible encoding functions of the form

$$m(x) = \tau(\phi(x)),$$

where the "template function" $\tau(\phi)$ is some given continuously increasing function with a bounded range, and $\phi(x)$ is any affine transformation of the form (A.6) with $\beta > 0$. Taking the template function as given, this defines a two-parameter family of possibilities (with parameters α and β).

For example, suppose that the template function is given by $\tau(y) = \underline{m}$ for all $y \le 0$, $\tau(y) = \underline{m} + (\bar{m} - \underline{m}) \cdot y$ for all $0 \le y \le 1$, and $\tau(y) = \bar{m}$ for all $y \ge 1$. Then, the two-parameter family of possible encoding functions consists of the family of affine functions, truncated so that the value of $m(x)$ remains within the bounds $[\underline{m}, \bar{m}]$. This is the class of alternative encoding functions assumed to be possible in models of context-sensitive encoding like that of Soltani et al. (2012), as discussed further below. However, one might well consider other template functions. The adaptation shown in Fig. 2 is an example of range adaptation in which the template function has a sigmoid shape (an affine transformation of the normal CDF).

Let us use the notation $\mathcal{F}(\tau)$ for the two-parameter family of possible encoding rules associated with a given template function τ. Then, the hypothesis of *fast range adaptation* assumes that in each environment, if magnitudes in that environment are drawn from a prior with density function $f(x)$, the encoding function $m(x)$ used in

[36] We can also establish a version of the Corollary for the cases in which the objective depends on $u(x)$ rather than x itself; in this case, the affine transformation must be an affine transformation of the utility scale rather than an affine transformation of the financial return. We omit the details, which are relatively straightforward.

that environment will be chosen from among the parametric family $\mathcal{F}(\tau)$ so as to minimize P^e, where (A.2) is evaluated using the environment-specific density $f(x)$. That is, this hypothesis requires only that the parameters α and β of the transformation $\phi(x)$ be optimized for each environment; the template function $\tau(\phi)$ is assumed to remain unchanged (or to adjust only over some much lower frequency).

This is a weaker optimality criterion than the one assumed in the theory of optimal adaptation presented above since it is only required that $m(x)$ be optimal within the restricted family $\mathcal{F}(\tau)$, rather than that it be optimal within some much broader class of functions, such as the class of all nondecreasing functions with a given bounded range. Nonetheless, even this weaker optimality criterion suffices to deliver the result of the Corollary.

Proposition 3. Let $F(x)$ and $\tilde{F}(x)$ be the CDFs corresponding to two possible prior distributions for the magnitude x, such that there exists an increasing affine transformation $\tilde{\phi}(x)$ with the property that $\tilde{F}(x) = F(\tilde{\phi}(x))$ for all x. And, suppose that $m(x)$ and $\tilde{m}(x)$ are the optimal encoding functions in the case of priors $F(x)$ and $\tilde{F}(x)$, respectively, from within the parametric family of encoding functions $\mathcal{F}(\tau)$ associated with some template function $\tau(\phi)$. Then, regardless of the template function used to define the parametric family, it must be the case that $\tilde{m}(x) = m(\tilde{\phi}(x))$ for all x. That is, $\tilde{m}(x)$ can be obtained from $m(x)$ by shifting and stretching the state space in the same way as is required to obtain $\tilde{F}(x)$ from $F(x)$.

Proof. Just as in the proof of Proposition 1, the value of P^e depends only on the normalized encoding function $n(q)$ implied by a given encoding function (which in the present context means by a given affine transformation) and a given prior; the normalized encoding function implied by a given prior $F(x)$ and affine transformation $\phi(x)$ is defined as

$$n(q) \equiv \frac{\tau(\phi(F^{-1}(q))) - \underline{m}}{\bar{m} - \underline{m}}. \tag{B.1}$$

Let $\bar{n}(q)$ be the normalized encoding function that minimizes P^e (for the given template function $\bar{m}(x)$).

Now, suppose that in the environment with prior $F(x)$, the optimal encoding function within the parametric family is given by $m(x) = \tau(\phi(x))$. The corresponding optimal normalized encoding function $\bar{n}(q)$ is then given by the right-hand side of (B.1). Since the assumption that $\tilde{F}(x) = F(\tilde{\phi}(x))$ implies that $F^{-1}(q) = \tilde{\phi}(\tilde{F}^{-1}(q))$, we must also have

$$\bar{n}(q) = \frac{\tau\left(\phi\left(\tilde{\phi}\left(\tilde{F}^{-1}(q)\right)\right)\right) - \underline{m}}{\bar{m} - \underline{m}}.$$

This implies that, in the environment with prior $\tilde{F}(x)$, the affine transformation $\phi(\tilde{\phi}(x))$ results in a normalized encoding function $\bar{n}(q)$, and hence that this affine transformation minimizes P^e. Hence, the optimal encoding function for this environment will be

$$\tilde{m}(x) = \tau(\phi(\tilde{\phi}(x))) = m(\tilde{\phi}(x)),$$

as the Proposition asserts. □

Thus, the weaker hypothesis of fast range adaptation in each of the environments encountered by a decision maker suffices to imply that, if the prior shifts from the distribution $f(x)$ shown in Fig. 2 to the alternative distribution $\tilde{f}(x)$ (a horizontal translation of $f(x)$), and the optimal encoding function for the first environment is given by the function $m(x)$ shown in the figure, then the optimal encoding function for the second environment will be the horizontal translation $\tilde{m}(x)$, as shown. The fact that this prediction (the basis for our experimental test) requires only a hypothesis of optimal range adaptation makes it more plausible that adaptation of the kind required can occur relatively rapidly. For the adaptation in question requires only estimation of optimal values for the two parameters α and β, and even a relatively small number of observations of a new environment can suffice to allow at least a rough estimate of the way in which these two parameters should be adjusted.

B2. Comparison with the literature

Examples of optimal range adaptation models of this kind include (Rustichini et al., 2017; Khaw et al., 2020; Zhang et al., 2020). Note that the assumed form of the template function is different in each of these cases. For example, Rustichini et al. (2017) assume a truncated linear template function. Khaw et al. (2020) instead propose a theory of optimal range adaptation in which the template function is a logarithmic transformation of a kind often assumed in sensory domains to which "Weber's Law" applies. Zhang et al. (2020) assume a template function given by a truncated log-odds transformation. This illustrates the need for a theory of the determination of the template function of the kind that we propose below (Section 5.2.3).

Optimal range adaptation in the sense proposed above represents a generalization of a more common conception of range adaptation (dating at least to the work of Parducci, 1965), in which it is assumed that the frequency distribution $f(x)$ has finite support $[\underline{x}, \bar{x}]$ (the "range" of stimulus magnitude). In this case, the distribution of internal representations r associated with a given stimulus x depends only on its "range adapted" value $\hat{x} \equiv (x - \underline{x})/(\bar{x} - \underline{x})$, indicating where x lies relative to the boundaries of the range. This idea corresponds to a truncated-linear template function, together with the hypothesis that the parameters α and β are chosen, for any given frequency distribution $f(x)$, so that the affine function $\phi(x)$ maps the stimulus range $[\underline{x}, \bar{x}]$ into the interval $[\underline{m}, \bar{m}]$.

The difference between our theory and this classic formulation is that we do not assume that the template function must be of the truncated-linear form. Instead, the shift in the encoding function shown in Fig. 2 can be viewed as an example of fast range adaptation only on the assumption that the template function has a sigmoid shape, and the sigmoid shape is important for our prediction of outlier blindness. It is because of the sigmoid shape that the encoding function is less steeply sloped between the points x_1 and x_2 when those points are in a tail of the prior distribution, rather than when they are near the prior mode.[37]

[37] Note, however, that under the assumption of a truncated-linear template function, as assumed for example in Soltani et al. (2012), outlier

In our model, the template function is endogenously determined as an optimal adaptation to the agent's environment within a much more flexible family, but only through an adaptive process that occurs much more slowly. In the short run, the template function is fixed, and only the parameters α and β of the normalization adjust to the current frequency distribution of magnitudes to be encoded; but, eventually, the template function adjusts as well. Thus, the template function is also assumed to be optimized to maximize the expected probability of a correct choice, but with the expectation calculated by averaging over all of the different short-run contexts that are encountered over the long run.

B3. Long-run adaptation of the template function

In our complete model, adaptation of the encoding rule occurs on two separate time scales. Given the current template function $\tau(\phi)$, the affine transformation $\phi(x)$ adjusts relatively quickly in a way that is optimal for the current statistics of the environment. But the template function is also adjusted in a way that adapts it to the statistics of the magnitudes that are encountered, only at a much lower frequency. Thus, $\tau(\phi)$ should be optimal relative to the statistics of the decision maker's environment as well, but, because $\tau(\phi)$ adapts only slowly, the relevant statistics are average frequencies over a fairly long run, while $\phi(x)$ adjusts more rapidly and so is at each point in time optimal (or nearly optimal) relative to the frequencies with which different magnitudes are encountered in a particular, more transitory environment.

To formalize this idea, we extend the theory of efficient coding presented above in the following way. Suppose that a decision maker transitions among environments (indexed by a parameter θ), with the distribution of magnitudes in each environment given by a distribution $F(x|\theta)$. Each short-run environment lasts long enough for range adaptation to that environment to occur. Thus, there is an affine transformation $\phi(x|\theta)$ for each possible θ that is adapted to the distribution $F(x|\theta)$. Finally, suppose that, over the long run, different environments θ are encountered with frequencies corresponding to a distribution $p(\theta)$. We assume that the template function adjusts slowly enough that it is the same in each of the environments, but that it is optimized with respect to the long-run frequency distribution $p(\theta)$.

For any environment, let $P^e(F, \phi, \tau)$ be the value of the objective (A.4) for that environment, if magnitudes are drawn from the distribution $F(x)$, and encoded using an affine transformation of the state space $\phi(x)$ and a template function $\tau(\phi)$, and the normalized encoding function is $n(q)$ as defined in (B.1). The complete theory of efficient coding with two time scales can then be expressed as follows. There is assumed to be a single template function $\tau(\phi)$, and an affine transformation $\phi(x|\theta)$ for each possible environment θ, such that (i) for each environment θ, $\phi(x|\theta)$ minimizes

$$P^e(F(\cdot|\theta),\ \phi(\cdot|\theta),\ \tau),$$

given the distribution of magnitudes associated with that environment, and (ii) $\tau(\phi)$ minimizes

$$\int P^e(F(\cdot|\theta),\ \phi(\cdot|\theta),\ \tau)\, p(\theta) d\theta$$

from among the class of all nondecreasing functions with range $[\underline{m}, \bar{m}]$.

Condition (ii) of this definition determines the shape of the optimal template function. Under some (not implausible) conditions, a sigmoid shape of the kind assumed in Fig. 2 will be optimal. For example, consider the case in which, for each possible environment θ, $F(x|\theta)$ is a Gaussian distribution $N(\mu_\theta, \omega_\theta^2)$, where the mean and standard deviation depend on the environment. In this case, each of the priors can be obtained from any of the others through an affine transformation of the state space. It then follows from Proposition 3 that the optimal encoding function for each environment can be obtained from the function for some single environment, through this same affine transformation.

Thus, the optimal encoding function for each environment θ must be of the form

$$m(x|\theta) \ = \ \tau\left(\frac{x-\mu_\theta}{\omega_\theta}\right)$$

for some nondecreasing function $\tau(z)$ that is the same for all θ. It follows from this that P^e will have the same value for each environment θ, equal to the value of P^e in the case in which $F(x)$ is a standard normal distribution, and $\phi(x)$ is the identity mapping ($\phi(x) = x$ for all x). Condition (ii) then reduces to the requirement that $\tau(z)$ be chosen so as to minimize P^e (as defined in (A.2) for the case of a prior over z that is a standard normal distribution). This is the kind of efficient coding problem discussed in the previous section. The asymptotic result (A.13) implies that, at least in the small-noise limit, the optimal template function $\tau(z)$ will be a sigmoid function (in fact, approaching the function $\Phi(z)$ as the noise is made small).

This provides another justification for an encoding function with a sigmoid shape, as shown in the lower right panel of Fig. 2, and hence for the prediction that we test in our experiment. Thus, the prediction that we test can be obtained as a prediction of either of two possible theories of efficient coding. The first theory is one in which the encoding function $m(x|\theta)$ is fully optimal for the distribution of magnitudes $F(x|\theta)$ for each environment θ; since this theory implies range adaptation (see the Corollary above), a change in the prior of the kind shown in the figure should result in a shift of the encoding function of the kind that is shown. The second possible theory is one in which only the affine transformation $\phi(x|\theta)$ is optimized for each environment, while the template function remains the same across environments. On the assumption that the long-run optimal template function (i.e., the one that is optimal for the full range of environments encountered over the long run) has a sigmoid shape, this theory

blindness would also be observed: there should be no ability to discriminate between stimuli outside the range beyond which the encoding function is truncated. And, in the case of a prior distribution with long tails (say, a Gaussian distribution), the tails of the distribution will necessarily be outside the finite range over which the encoding function is increasing.

will also imply that a shift in the distribution of magnitudes of the kind shown in the figure should result in a shift in the encoding function of the kind shown in the figure.

One might wonder why we bother to present this more complex version of our theory, given that the basic version of our model already provides a reason for outlier blindness to be observed. But the conclusion that adaptation occurs relatively quickly is important for financial applications of our theory, inasmuch as investment contexts are often not fixed over time but shifting.

If the encoding rule adapts only to the frequency distribution of monetary amounts encountered over some very long run, one might suppose that this distribution will be quite dispersed, so that none of the values that can be encountered in the current short-run context would be much more unlikely than any of the others. In such a case, one might suppose that none of the values that are encountered should really be "outliers" relative to the range of values to which perception has been adapted, and that the accuracy with which different magnitudes can be discriminated should be approximately constant. If, instead, there is fast range adaptation to the current short-run frequency distribution, it is easier to believe that some values that can occur with non trivial probability are nonetheless far enough from the center of the short-run prior to be more poorly distinguished than values nearer the mode of that distribution. Thus, it is in this case that imprecision in the internal representation of monetary amounts is likely to lead not just to noise in subjective valuations, but also to systematic bias (i.e., outlier blindness).

Appendix C. Optimality of conservative bias

Thus far, we have analyzed only optimal adaptation of the encoding function $m(x)$; these results suffice to explain the predictions of our model for the experiments discussed in Section 2. However, our discussion of the implications of our theory for tail risk neglect in Section 3.3 invokes further results about the optimal "decoding" of the internal representation r to produce a subjective estimate $\hat{x}(r)$ of the magnitude that has been observed. We turn now to this issue.

C1. Bayesian decoding

Suppose that an investor knows (either on the basis of past experience, or description of the possible outcomes) that a given investment will yield one or another of a set of possible returns $\{x_s\}$, and that each of the possible returns x_s (associated with some state of the world s) is separately encoded, resulting in an internal representation $r_s \sim N(m(x_s), \sigma^2)$. To choose the investment with the highest expected return, the investor's decision rule will be a function of the conditional means $\hat{x}_s \equiv \mathrm{E}[x_s \,|r_s]$ for each of the possible states of the world. Specifically, the investor should choose the investment for which the weighted average of the estimates $\hat{x}_s$ is highest, where the weights applied to the different states are given by the investor's subjective estimate of the probability of each of these outcomes.[38]

Thus, an optimal decision rule requires that the internal representation be "decoded" to obtain an estimate $\hat{x}_s$ of the return associated with each state, where the optimal decoding rule associates with an internal representation r the implied conditional mean $\mathrm{E}[x\,|r]$. This is often called "Bayesian decoding" since the estimate $\hat{x}(r)$ is the mean of the Bayesian posterior distribution for x if the prior distribution is updated using Bayes's rule following observation of the noisy representation r.

If the prior distribution is specified by a density function $f(x)$ and the encoding function is $m(x)$, then the Bayesian posterior mean estimate implied by any internal representation r is given by

$$\hat{x}(r) \;=\; \frac{\int \exp\left\{-\frac{1}{2}\left(\frac{r-m(x)}{\sigma}\right)^2\right\} x f(x)\,dx}{\int \exp\left\{-\frac{1}{2}\left(\frac{r-m(x)}{\sigma}\right)^2\right\} f(x)\,dx}. \tag{C.1}$$

Here we assume that $f(x)$ integrates to 1 (as required in order for it to represent a probability distribution), and that $f(x)$ has thin enough tails for the integrals

$$\int_{-\infty}^{0} x f(x)\,dx, \qquad \int_{0}^{\infty} x f(x)\,dx$$

both to be finite (so that the prior distribution has a first moment); then both the numerator and denominator of (C.1) are well-defined.

We can show quite generally that Bayesian decoding implies that the estimate $\hat{x}$ will not be unbiased and, more specifically, that there will be an overall conservative bias. If we define the bias in the case of a true magnitude x as $b(x) \equiv \mathrm{E}[\hat{x}\,|x]$, then it must be the case that

$$\begin{aligned}\mathrm{cov}(b(x),\,x) &= \mathrm{E}[b(x)\cdot x] \;=\; \mathrm{E}[\mathrm{E}[b(x)\cdot x\,|r]]\\ &= \mathrm{E}[\mathrm{E}[\mathrm{E}[x\,|r]\cdot x \,-\, x^2\,|r]]\\ &= \mathrm{E}[\mathrm{E}[x\,|r]^2 \,-\, \mathrm{E}[x^2\,|r]]\\ &= -\mathrm{E}[\mathrm{var}(x\,|r)] \;\le\; 0.\end{aligned}$$

Moreover, the final inequality must be strict unless perfectly precise identification of the true magnitude is possible with probability one. Thus, the bias must be non-zero with positive probability, and it must be negatively correlated with the true magnitude (conservatism).

This result holds quite generally, as a simple consequence of noisy encoding combined with Bayesian decoding. The kind of relationship between the average estimate $\mathrm{E}[\hat{x}\,|x]$ and the true magnitude x that is implied by an efficient coding rule is illustrated by Fig. 8 in the main text for the case of a Gaussian prior distribution of returns.

C2. Biased probability weights in the case of decisions from description

In the discussion above, we have supposed that the magnitude x that must be encoded with noise and then

[38] This will be true regardless of whether these weights are formed by sampling from memory, in the case of a decision from experience, or the weights are derived from a noisy internal representation of the probabilities that the decision maker has been told, in the case of a decision from description. The latter case is discussed further below.

decoded to make a decision is some monetary payoff or rate of return (whether observed or described). However, in the case of so-called "decisions from description," the probability information that is presented to the decision maker must also be encoded and then decoded to evaluate the gamble. The finiteness of the decision maker's cognitive resources implies that information about stated (or calculated) probabilities should also be encoded with noise, so that the problem of optimal decoding of the noisy internal representation arises in the case of information about probabilities as well.

Consider, for example, the case of choice between simple gambles, each of which offers some payoff x with probability p, and zero otherwise. The characteristics of each gamble are then specified by the two quantities x and p. If we suppose that x and p are separately represented, by internal representations r_x and r_p, respectively, and, furthermore, that under the prior distribution x and p are distributed independently of one another, then the decision rule that maximizes the decision maker's expected financial gain from her decision will choose the gamble with the highest value of $\hat{p}\cdot\hat{x}$, where $\hat{p}(r_p)\equiv \mathrm{E}[p\,|r_p]$ and $\hat{x}\equiv \mathrm{E}[x\,|r_x]$.[39] Thus, optimal decoding of the noisy internal representation of the probability information will again correspond to Bayesian decoding.

It then follows from the discussion above that the estimate $\hat{p}$ will be biased, and, furthermore, that the bias $b(p)$ will satisfy

$$\mathrm{cov}(b(p),\, p) \;=\; -\mathrm{E}[\mathrm{var}(p\,|r_p)] \;<\; 0.$$

Thus, our theory predicts that the probability weight used in choosing between gambles will differ on average from the true probability of the gamble paying off, and the bias will be conservative (a tendency to underestimate larger probabilities and to overestimate smaller ones).

Now, let us assume a more specific model of the noisy encoding of probability information,

$$r_p \;\sim\; N(\tilde{m}(p),\, \sigma_p^2), \tag{C.2}$$

by analogy with our model of the encoding of monetary payoffs, where $\tilde{m}(p)$ is an increasing function that must satisfy the bounds

$$\underline{m} \;\le\; \tilde{m}(p) \;\le\; \bar{m} \tag{C.3}$$

for all $0<p<1$. Then, we can obtain a stronger result about the degree to which small probabilities will be overestimated on average using (C.1). Note that in our theory, the function

$$w(p) \;\equiv\; \mathrm{E}[\hat{p}\,|p]$$

plays a role analogous to the probability weighting function in Prospect Theory (Kahneman and Tversky, 1979), indicating the average weight that is placed on an outcome with true probability p. It is thus useful to study the behavior of this function for small values of p.

To apply (C.1) to the case of probability estimation, we replace the variable x in the formula by p, the standard deviation σ by σ_p, and so on. We then observe that the exponential factor appearing in this formula must satisfy the lower bound

$$\exp\left\{-\frac{1}{2}\left(\frac{r_p-\tilde{m}(p)}{\sigma_p}\right)^2\right\} \ge \min\left\{\exp\left(-\frac{1}{2}\left(\frac{r_p-\underline{m}}{\sigma_p}\right)^2\right),\; \exp\left(-\frac{1}{2}\left(\frac{r_p-\bar{m}}{\sigma_p}\right)^2\right)\right\}$$

for any real number r. Substituting this expression for the exponential factor in the numerator of (C.1), we obtain a positive lower bound for the numerator.

The exponential factor must also be bounded above by 1 for all r; hence, the denominator of (C.1) can be no greater than 1. The lower bound for the numerator and the upper bound for the denominator can then be used to establish the bound

$$\hat{p}(r_p) \;\ge\; \bar{p}\cdot\exp\left\{-\frac{1}{2}\left(\frac{r_p-\bar{m}}{\sigma_p}\right)^2\right\}$$

for all $r_p \le m^* \equiv (\underline{m}+\bar{m})/2$, and the bound

$$\hat{p}(r_p) \;\ge\; \bar{p}\cdot\exp\left\{-\frac{1}{2}\left(\frac{r_p-\underline{m}}{\sigma_p}\right)^2\right\}$$

for all $r_p \ge m^*$, where $\bar{p}\equiv \mathrm{E}[p]$ is the prior mean for the probability p.

We can then compute a lower bound for the average estimate $\hat{p}$, conditional on any true probability p, by integrating this bound for $\hat{p}(r_p)$ over the distribution (C.2) of possible values for r_p. We obtain the bound

$$\begin{aligned}
\mathrm{E}[\hat{p}\,|p] \ge{}& \int_{-\infty}^{m^*} \frac{1}{\sqrt{2\pi}\sigma_p}\exp\left(-\frac{1}{2}\left(\frac{r_p-\tilde{m}(x)}{\sigma_p}\right)^2\right) \\
&\cdot\bar{p}\cdot\exp\left(-\frac{1}{2}\left(\frac{r_p-\bar{m}}{\sigma_p}\right)^2\right)dr_p \\
&+\int_{m^*}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_p}\exp\left(-\frac{1}{2}\left(\frac{r_p-\tilde{m}(x)}{\sigma_p}\right)^2\right) \\
&\cdot\bar{p}\cdot\exp\left(-\frac{1}{2}\left(\frac{r_p-\underline{m}}{\sigma_p}\right)^2\right)dr_p \\
={}& \frac{\bar{p}}{\sqrt{2}}\left\{\exp\left(-\frac{1}{2}\bar{z}(p)^2\right)\Phi(-\underline{z}(p)) \right. \\
&\left. +\exp\left(-\frac{1}{2}\underline{z}(p)^2\right)\Phi(-\bar{z}(p))\right\} \\
\equiv{}& \underline{w}(p),
\end{aligned}$$

where we define

$$\underline{z}(p) \;\equiv\; \frac{\tilde{m}(p)-\underline{m}}{\sqrt{2}\sigma_p}, \qquad \bar{z}(p) \;\equiv\; \frac{\bar{m}-\tilde{m}(p)}{\sqrt{2}\sigma_p}.$$

The lower bound $\underline{w}(p)$ is easily seen to be strictly positive for all $\tilde{m}(p)$ satisfying the bounds (C.3). Furthermore, we see that $\underline{w}(p)$ remains bounded away from zero as $\tilde{m}(p)$ approaches its lower bound $\underline{m}$. Since this is a lower bound, it follows that $w(p)$ is also bounded away from zero as p approaches zero. Hence, for all small enough

[39] See (Khaw et al., 2020) for further explanation of this model of choice under risk in the case of decisions from description.

probabilities $p > 0$, the average decoded value $\hat{p}(r_p)$ will necessarily be larger than p on average. Yet, the same is not true for all probabilities. Indeed, using a similar argument, we can show that, for all probabilities $p < 1$ close enough to 1, the average decoded value will necessarily be *less* than p on average. Thus, our model predicts overestimation of low probabilities and over-estimation of large probabilities.

It follows that, in the case of decisions from description, we should expect low-probability gambles to be evaluated as if the probability of the gamble paying off were greater than it really is, fairly often. In the case of a gamble that combines a low positive value of p with a value for x relatively far in the right tail of the prior distribution, we should expect $\hat{x}$ to be less than x on average (the conservative bias discussed above), but, at the same time, we should expect $\hat{p}$ to be larger than p on average. It can then easily be the case that the subjective estimate $\hat{p} \cdot \hat{x}$ of the expected value of the gamble exceeds its true expected value more often than not. It is, furthermore, quite possible that the subjective estimate of the expected value of the gamble will exceed the subjective estimate of the value of receiving a payment of px with certainty, given that there is noopportunity for probability distortion in the latter case. Thus, one could observe apparently risk-seeking behavior, despite under-estimation of the magnitude of right-tail payoffs.[40]

References

Aridor, G., Grechi, F., Woodford, M., 2020. Adaptive efficient coding: a variational autoencoder approach. pre-print posted on bioRXiv, https://www.biorxiv.org/content/10.1101/2020.05.29.124453v1.
Ariely, D., 2010. Predictably Irrational: The Hidden Forces That Shape Our Decisions. HarperCollins, New York.
Barberis, N., 2011. A model of casino gambling. Manag. Sci. 58, 35–51.
Barberis, N.C., 2013. The psychology of tail events: progress and challenges. Am. Econ. Rev. Papers Proc. 103, 611–616.
Barberis, N.C., Huang, M., 2008. Stocks as lotteries: the implications of probability weighting for security prices. Am. Econ. Rev. 98, 2066–2100.
Barseghyan, L., Molinari, F., O'Donoghue, T., Teitelbaum, J.C., 2013. The nature of risk preferences: evidence from insurance choices. Am. Econ. Rev. 103, 2499–2529.
Benabou, R., 2015. The economics of motivated beliefs. Rev. Econ. Polit. 125, 665–685.
Bhui, R., Gershman, S.J., 2018. Decision by sampling implements efficient coding of psychoeconomic functions. Psychol. Rev. 125, 985–1001.
Bordalo, P., Gennaioli, N., Shleifer, A., 2012. Salience theory of choice under risk. Q. J. Econ. 127, 1243–1285.
Bredenberg, C., Simoncelli, E. P., Savin, C., 2020. Learning efficient task-dependent representations with synaptic plasticity. pre-print posted on bioRXiv, https://www.biorxiv.org/content/10.1101/2020.06.19.162172v1.full.pdf.
Carandini, M., Heeger, D.J., 2012. Normalization as a canonical neural computation. Nat. Rev. Neurosci. 13, 51–62.
Cheng, I.H., Raina, S., Xiong, W., 2014. Wall street and the housing bubble. Am. Econ. Rev. 104, 2797–2829.
Dixit, A.K., Nalebuff, B.J., 1991. Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life. W. W. Norton & Company, New York.
Donnelly, C., Embrechts, P., 2010. The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. ASTIN Bull. 40, 1–33.

[40] For similar reasons, of course, imprecise probability estimates can easily lead to more risk-averse choices in the case of gambles with negative tail risk. This is indeed observed in the case of decisions from description (e.g., the experiments of Kahneman and Tversky), though not for decisions from experience.

Erev, I., Ert, E., Roth, A.E., Haruvy, E., Herzog, S.M., Hau, R., Hertwig, R., Stewart, T., West, R., Lebiere, C., 2010. A choice prediction competition: choices from experience and from description. J. Behav. Decis. Mak. 23, 15–47.
Fama, E.F., 1965. The behavior of stock market prices. J. Bus. 38, 34–105.
Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Meth. 39, 175–191.
Frydman, C., Jin, L.J., 2020. Efficient coding and risky choice. Working Paper.
Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H.E., 2003. A theory of power-law distributions in financial market fluctuations. Nature 423, 267–270.
Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H.E., 2006. Institutional investors and stock market volatility. Q. J. Econ. 121, 461–504.
Gabaix, X., Laibson, D., 2017. Myopia and discounting. NBER Working Paper No. 23254.
Ganguli, D., Simoncelli, E. P., 2016. Neural and perceptual signatures of efficient sensory coding. Working PaperarXiv:1603.00058v1
Gennaioli, N., Shleifer, A., 2018. A crisis of Beliefs. Investor Psychology and Financial fzragiity. Princeton University Press, Princeton.
Gennaioli, N., Shleifer, A., Vishny, R., 2012. Neglected risks, financial innovation, and financial fragility. J. Financ. Econ. 104, 452–468.
Gennaioli, N., Shleifer, A., Vishny, R., 2015. Neglected risks: the psychology of financial crises. Am. Econ. Rev. 105, 310–314.
Giglio, S., Maggiori, M., Stroebel, J., Utkus, S., 2020. Inside the mind of a stock market crash. Working Paper.
Girshick, A.R., Landy, M.S., Simoncelli, E.P., 2011. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat. Neurosci. 14, 926–932.
Glimcher, P., 2014. Understanding the hows and whys of decision-making: from expected utility to divisive normalization. Cold Spring Harb. Symp. Quant. Biol. 79, 169–176.
Hertwig, R., Barron, G., Weber, E.U., Erev, I., 2004. Decisions from experience and the effect of rare events in risky choices. Psychol. Sci. 15, 534–539.
Kahneman, D., 2003. Maps of bounded rationality: psychology for behavioral economics. Am. Econ. Rev. 93, 1449–1475.
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–291.
Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financ. Stud. 27, 2841–2871.
Khaw, M.W., Glimcher, P.W., Louie, K., 2017. Normalized value coding explains dynamic adaptation in the human valuation process. Proc. Natl. Acad. Sci. 114, 12696–12701.
Khaw, M.W., Li, Z., Woodford, M., 2020. Cognitive imprecision and small--stakes risk aversion. Rev. Econ. Stud.
Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121, 1133–1165.
Koszegi, B., Szeidl, A., 2013. A model of focusing in economic choice. Q. J. Econ. 128, 53–104.
Kumar, A., 2009. Who gambles in the stock market? J. Finance LXIV, 1889–1933.
Kumar, A., Page, J., Spalt, O., 2011. Religious beliefs, gambling attitudes, and financial market outcomes. J. Financ. Econ. 102, 671–708.
Laughlin, S.B., 1981. A simple coding procedure enhances a neuron's information capacity. Z. Nat. C 36, 910–912.
Lo, A., 2017. Adaptive Markets: Financial Evolution at the Speed of Thought. Princeton University Press, Princeton.
Lochstoer, L.A., Muir, T., 2021. Volatility expectations and returns. NBER Working Paper No. w28102.
Mandelbrot, B.B., 1957. Fractales, Hasard et Finance. Flammarion, Paris.
McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? J. Finance 71, 5–32.
Minsky, H.P., 1986. Stabilizing an Unstable Economy. MGraw-Hill, New York.
Padoa-Schioppa, C., Assad, J.A., 2008. The representation of value in the orbitofrontal cortex is invariant for changes of menu. Nat. Neurosci. 11, 95–102.
Padoa-Schioppa, C., Rustichini, A., 2014. Rational attenion and adaptive coding: a puzzle and a solution. Am. Econ. Rev. 104, 507–513.
Parducci, A., 1965. Category judgment: a range-frequency model. Psychol. Rev. 72, 407–418.
Payzan-LeNestour, E., 2018. Can people learn about 'black swans'? experimental evidence. Rev. Financ. Stud. 31, 4815–4862.
Payzan-LeNestour, E., Pradier, L., Putnins, T.J., 2021. The "waterfall illusion" in financial markets: how risk perception is distorted after exposure to extreme risk. Working Paper.
Polania, R., Woodford, M., Ruff, C.C., 2019. Efficient coding of subjective value. Nat. Neurosci. 22, 134–142.

Rabin, M., 2013. An approach to incorporating psychology into economics. Am. Econ. Rev. 103, 617–622.
Robson, A. J., Whitehead, L. A., 2019. Adaptive cardinal utility. unpublished paper.
Rustichini, A., Conen, K.E., Cai, X., Padoa-Schioppa, C., 2017. Optimal coding and neuronal adaptation in economic decisions. Nat. Commun. 8 (1208).
Soltani, A., Martino, B.D., Camerer, C., 2012. A range-normalization model of context-dependent choice: a new model and evidence. PLoS Comput. Biol. 8. doi:10.1371/journal.pcbi.1002607.
Steiner, J., Stewart, C., 2016. Perceiving prospects properly. Am. Econ. Rev. 106, 1601–1631.
Summerfield, C., Tsetsos, K., 2015. Do humans make good decisions? Trends Cogn. Sci. 19, 27–34.
Taleb, N.N., 2004. Fooled By Randomness: The Hidden Role of Chance in Life and in the Markets. Penguin Books, London.
Thurstone, L.L., 1927. A law of comparative judgment. Psychol. Rev. 34, 273–286.
Tobler, P.N., Fiorillo, C.D., Schultz, W., 2005. Adaptive coding of reward value by dopamine neurons. Science 307, 1642–1645.
Tremblay, L., Schultz, W., 1999. Relative reward preference in primate orbitofrontal cortex. Nature 398, 704–708.
Ungemach, C., Chater, N., Stewart, N., 2009. Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? Psychol. Sci. 20, 473–479.
Wei, X.X., Stocker, A.A., 2015. A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. Nat. Neurosci. 18, 1509–1517.
Wei, X.X., Stocker, A.A., 2017. Lawful relation between perceptual bias and discriminability. Proc. Natl Acad. Sci. 114, 10244–10249.
Woodford, M., 2012. Inattentive Valuation and Reference-Dependent Choice. Working Paper. Columbia University.
Woodford, M., 2012. Prospect theory as efficient perceptual distortion. Am. Econ. Rev. 102, 41–46.
Woodford, M., 2020. Modeling imprecision in perception, valuation and choice. Ann. Rev. Econ. 12, 579–601.
Zhang, H., Ren, X., Maloney, L.T., 2020. The bounded rationality of probability distortion. Proc. Natl. Acad. Sci. 36, 22024–22034.