"Outlier Blindness": Efficient Coding Generates an Inability to Represent Extreme Values*

By Elise Payzan-LeNestour and Michael Woodford†

Draft: April 1, 2019

How do economic agents perceive extreme values? Building on a well-established theory from neuroscience, we hypothesize that economic agents are inherently hampered in the way they perceive extreme values because the human brain evolved to devote neural activity to representing the most probable values at the expense of the improbable ones. We find support for this hypothesis in a series of controlled laboratory experiments. (JEL: C91, D87)

Keywords: Neuroeconomics, Experiments, Tail risk, Instability, Efficient coding, Adaptation, Decision making under uncertainty.

* We are indebted to Peter Bossaerts, Philippe Gouillou, Lawrence Jin, Carsten Murawski, Amos Nadler, Antonio Rangel, Agnieszka Tymula, seminar participants at Caltech, Melbourne University, the Behavioral and Experimental Economics and Finance workshop at the University of Sydney, 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 Peter Dayan for providing us with extremely detailed comments on a preliminary draft. The analyses and conclusions set forth are those of the authors. Daniela Grinblatt and Philippe Gouillou provided outstanding research assistance. Funding from Columbia University and UNSW Business School is gratefully acknowledged.

† Payzan-LeNestour: UNSW Business School, University of New South Wales, Sydney, elise@unsw.edu.au. Woodford: Department of Economics, Columbia University, New York, michael.woodford@columbia.edu.
**Introduction**

It has long been established that “tail risk” (the fact that asset returns are fat-tailed) is an essential feature of financial markets, and the capacity to accurately represent it is an important factor in financial decision making.\(^1\) Yet, the study of how individuals actually represent tail risk is fairly recent. Evidence suggests that people doubly misrepresent tail risk, with significant consequences for financial stability: they under-estimate extreme events until they occur (‘ex ante neglect’) and assign them disproportionate weight once they occur (‘ex post overreaction’).\(^2\) In this paper, we set out to complete this picture by proposing the new idea that ex post overreaction does not occur unconditionally. Rather, it is specific to the kind of extreme outcomes that occur either as a prolonged series of shocks or after a series of warning signs (e.g., macroeconomic depressions). In contrast, extreme outcomes that come as a complete surprise (e.g., regime shift news or jumps), are underweighted when they occur.

In the new theory that we propose, the channel through which extreme events that are completely unexpected (henceforth, “outliers”) are underweighted is not preference-related or learning-related. Rather, it relates to perception (how the brain represents and processes information about the outside world), the very first step of the decision-making process on which learning hinges—beliefs are indeed updated on the basis of perceived rather than objective outcomes. We hypothesize that people misperceive the magnitude of outliers. Specifically, they perceive their magnitude as less extreme than in reality. Such ‘conservative bias’ originates from the way the human brain evolved to devote neural activity to representing the most probable outcomes at the expense of outliers. We document strong experimental evidence for this hypothesis, which builds on a well-established neuroscience theory on human perception called “efficient coding theory.”

The starting point of efficient coding theory is the observation that the representational capacities of the brain are limited (we have a finite number of neurons and a finite number of possible spike outputs of each neuron). Therefore, if a neuron’s limited outputs were allocated evenly to represent the potentially infinite number of possible values of a stimulus, then that neuron’s activity would allow for little if any discrimination between values, i.e., perception would be really coarse. The optimal solution to the problem

---

\(^1\)For evidence on the fat-tailedness of asset returns, see, e.g., Mandelbrot (1957), Fama (1965), Gabaix et al. (2003), Gabaix et al. (2006), Kelly and Jiang (2014).

\(^2\)For example, Gennaioli and Shleifer (2018).
Outlier Blindness

found by evolution—“efficient coding”—consists of ensuring that neurons learn which values are more likely to occur and allocate most of their spike outputs to representing the most probable values at the expense of the improbable values.\(^3\) This way of allocating neural activity implies that people are unable to properly represent values that are highly improbable relative to the range of values they are expecting (“outlier blindness”). This results in the perceived magnitude of outliers to be less extreme than the reality.\(^4\)

We examine whether outlier blindness may prevail in modern economic environments. Prima facie there is little reason to expect that it does, for two reasons related to payoff instability. First, a prerequisite for the occurrence of outlier blindness is that the agents hold expectations. What the agents expect is the range of values to which they have been repeatedly exposed—or, in the language of neuroscientists, “adapted.” Modern economic environments may be too unstable for adaptation to take place, owing to the frequent occurrence of regime shifts or jumps in values.\(^5\) Second, efficient coding works only for the type of setting in which it evolved, namely, the natural environment, which features unimodal stimulus distributions. It is however a ‘misfit’ in unstable settings which typically feature multimodal stimulus distributions. The human brain might be sufficiently flexible to adjust its coding method to match the statistics of the environment. For example, it might use a multimodal prior for the distribution of economic payoffs, in place of the unimodal prior used in efficient coding.

In Section I of this paper, we propose the novel idea (henceforth, “outlier blindness hypothesis”) that outlier blindness occurs in modern economic environments, despite the instability, inasmuch as i) adaptation occurs very quickly in the human brain, and ii) efficient coding is an ingrained principle that the brain employs irrespective of the statistics of the environment. Assumption i) aligns with recent theoretical proposal that efficient coding takes a particular form, called “range adaptation,” in which only two parameters of the frequency distribution associated with a given environment (for example, the mean and standard deviation) are to be updated when the environment shifts.\(^6\) This means that even a few observations in a new environment may suffice for adaptation to occur. There is a good deal of evidence for range adaptation and corollary high-frequency

\(^3\)For example, Laughlin (1981), Tobler et al. (2005), Wei and Stocker (2015), and Ganguli and Simoncelli (2016). Similar principles have been applied to models of imprecise perception in economic decision making by, e.g., Robson (2001) and Polania et al. (2019). For more references, see Section I.

\(^4\)For example, Woodford (2012), Polania et al. (2019), and Frydman and Jin (2018).

\(^5\)Prior field research has identified the presence of regime shifts or jumps as a major feature of the returns generating process. See Ang and Timmermann (2011) for a survey of the literature.

\(^6\)For example, Soltani et al. (2012) and Rustichini et al. (2017).
adaptation. Assumption ii) reflects the idea that viewed from evolutionary timescales, modern economic environments emerged yesterday and it is unlikely that the brain was reconfigured to match their statistics.

In Section II of this paper, we report on an experimental test of outlier blindness hypothesis and its key assumptions i) & ii). We ask university students to perform in the laboratory for significant amounts of money a perceptual task in which on each trial, they have to discriminate between two adjacent shades of grey. Shade intensity is drawn from a normal distribution whose mean and standard deviation change across blocks of 40 trials, to mimic the instability of real-world economic environments. Our motivation for using shades of grey rather than an economic stimulus (e.g., financial returns) is to circumvent the issue of separating purely perceptual factors from behavioral factors related to risk attitude (e.g., aversion to jump risk). Given our main purpose in this study is to isolate a purely perceptual phenomenon, disentangling the channel through which decision making is affected (preferences vs. beliefs) is crucial, but it is impossible in an economic decision-making task. To address the problem, we design a purely perceptual task involving a simple sensory stimulus. While doing so would generally mean going from bad to worse owing to the generalizability issue [the issue of whether one can apply findings obtained in one domain to another domain], it is not the case here inasmuch as efficient coding is stimulus invariant, i.e., evidence shows that it prevails similarly for the perception of a wide range of stimuli, from simple sensory stimuli (Schwartz and Dayan, 2007) to complex stimuli such as economic value (e.g., Tremblay and Schultz (1999), Padoa-Schioppa and Assad (2008), and Khaw et al. (2017)). What such stimulus invariance means is that one can learn about how individuals perceive economic payoffs from the findings obtained in purely perceptual tasks involving simple stimuli such as the one used here.

The core of our experimental strategy consists of assessing the accuracy of subject perception of outliers, which we define as shade values that fall at least three standard deviations away from the range of values to which the subject has been exposed—according to outlier blindness hypothesis, “adapted”—in the previous 40 trials, relative to subject accuracy in control test trials in which the subjects are presented with the exact same shade.

Footnotes:
7For example, Laughlin and Hardie (1978), Ohzawa et al. (1985), Fairhall et al. (2001), Bayer and Glimcher (2005), Tobler et al. (2005), Fairhall (2014).
8On the growing consensus in neuroscience that efficient coding is a universal neural principle and as such, it is expected to apply similarly across all kinds of stimuli, see, e.g., Carandini and Heeger (2012), Glimcher (2014), and Padoa-Schioppa and Rustichini (2014).
9Evidence suggests that this idea of a common principle underlying both purely perceptual and economic decision-making is not specific to efficient coding but applies to other psychological mechanisms as well. See, e.g., Frydman and Nave (2016).
values but this time they’re not outliers as they fall within the range of values to which the subject has been adapted in the previous 40 trials. This way we control for important confounds inherent in the presentation of extreme values. For instance, in light of the classical “Weber-Fechner’s law” of diminishing sensitivity (in which a change in stimulus value is harder to perceive if it occurs at a higher intensity level), one expects perceptual accuracy to be inherently decreased for extreme values simply because it is harder to discriminate between values located at the extremes on the scale of grey. Therefore, merely comparing subject accuracy when discriminating between extreme shade values to their mean accuracy when discriminating between the other possible shades of grey on the scale features an increased risk of Type I error. In contrast, our test, by controlling for the diminishing sensitivity phenomenon, is specifically designed to isolate the outlier blindness effect caused by efficient coding.

We find that perception accuracy is markedly decreased for outliers, as predicted by outlier blindness hypothesis. To strengthen the evidence for the latter and to rule out any possibility that such decrease reflect some experimental artefact, we run a placebo test in which the experimental task is the same as the original task except for the duration of the blocks of trials prior to the presentation of the outliers, which lasts for only 3 trials in the placebo test (versus 40 trials in the original experiment). So by design, in the placebo test adaptation does not occur (i.e., the brain does not have time to learn which values are more likely to be presented), and hence outlier blindness should not occur. Consistent with this prediction, the original effect vanishes in the placebo test.

We further find that increasing block duration to 5 trials is enough for the outlier blindness effect to reappear, which shows that adaptation occurs very quickly in humans, consistent with Assumption i) above. This finding suggests that all kinds of decision makers are susceptible to the outlier blindness bias, including those operating in the most unstable settings.

By changing stimulus distribution across blocks of trials, our task is meant to capture the instability of real-world financial markets which arises from the frequent occurrence of regime shifts and the like, both at short and long timescales. While the short timescale feature of our design (the imparted response time on each trial is 2 sec, see Section II) directly maps into contexts of short timescale economic decision-making such as intra-day trading, there is a growing consensus in neuroscience that common principles apply over both short and long timescales.\(^\text{10}\) In light of this, one expects the phenomenon of outlier

\(^{10}\text{See Brown et al. (2007) for a survey of the literature showing that many memory phenomena appear\)}}
blindness to prevail similarly for the perception of jumps at short (intra-day or daily) horizons and the perception of “black swans” (Taleb, 2007) in a macroeconomic context.

**Relation to the Economics Literature**

This paper relates to several strands of literature in economics. First, a growing literature in neuroeconomics seeks to understand the implications of imprecise perception (the idea that decisions are based on internal representations that are imprecise) for economic decision-making. For instance, Steiner and Stewart (2016) show how imprecise perception exhibits a key feature of Prospect Theory’s probability weighting function, namely, overweighting of small probability events and corresponding underweighting of high probability events. In the same vein, Gabaix and Laibson (2017) show how imprecise perception generates apparent temporal discounting in perfectly patient agents. Most closely related is recent work that attempts to understand the implications of efficient coding theory for economic decision-making. Padoa-Schioppa and Rustichini (2014) and Rustichini et al. (2017) note how the aforementioned range adaptation property generates inconsistency in economic choice if left uncorrected, and they show how range adaptation is corrected within neural decision circuits so that human behavior does not feature such inconsistency. Payzan-LeNestour et al. (2018) report findings consistent with the idea that efficient coding generates a perceptual bias (“after-effect”) in risk perception. Frydman and Jin (2018) show that efficient coding provides an explanation for the variation of risk taking across environments.

The current evidence for outlier blindness is not a foregone conclusion in several respects. First, outlier blindness hypothesis is a new theory whose core underlying assumptions (cf. (i) & ii) above) are to be put to the test. Second, even if outlier blindness hypothesis is true, it may prove impossible to collect enough evidence to demonstrate it because the outlier blindness effect runs in opposite direction to (and hence may be offset by) what one may consider the most standard effect associated with the perception of improbable values, namely the classic “oddball effect.” In the oddball effect, information processing of unexpected stimulus values is improved—and hence perception accuracy is increased—relative to the perception of expected values because the unexpected is salient similar over a wide range of timescales.

At first glance, the outlier blindness effect is reminiscent of the well-known phenomenon of “inattentiveness blindness” (Mack and Rock, 1998) in which the decision maker fails to notice the unexpected (see Simons and Chabris (1999) for a classic example). However the two biases are of different nature: outlier blindness is a perceptual bias generated by efficient coding whereas inattentiveness blindness is an attentional bias reflecting sparse attention (e.g., Sims (2003), Mackowiak and Wiederholt (2009), and Gabaix (2014)).
Outlier Blindness

(“attention-grabbing”) per se due to its novelty or special significance to the subject.\textsuperscript{12} The fact that we find evidence for outlier blindness shows that the outlier blindness effect dominates the oddball effect, suggesting the former is quite significant.

Another striking aspect of the present findings lies in the implications of outlier blindness for economic decision making. The underweighting of extreme events implied by outlier blindness (see Section I.A) contrasts with Prospect Theory (in which low-probability events are overweighted) and the solid empirical evidence for overweighting of extreme events once they occur. For example, Malmendier and Nagel (2011) show that experiencing a macroeconomic depression has long term impacts on risk aversion, and Gennaioli and Shleifer (2018) document how the news of Lehman collapse in 2008 triggered investor panic. Such ex post overweighting of extreme events is fully compatible with outlier blindness theory inasmuch as it concerns extreme events that occur either as a series of shocks (e.g., depressions) or within a series (e.g., the news of Lehman collapse, which came after the 2006 housing bubble burst).\textsuperscript{13} As such, they are not truly unexpected to the agents and hence do not qualify as outliers in the sense defined here. Outlier blindness hypothesis applies to the perception of outliers specifically. Thus, viewed together with the foregoing literature on ex post overweighting of extreme events, the present study suggests that individuals’ reaction to extreme events differs depending on the kind of extreme events under study.

As such, our study adds to a recent and growing literature on decision making under tail risk. Studies on “neglected risks” provide profound insights into how agents underestimate the probability of extreme events until they occur as a result of both the availability heuristic (e.g., Gennaioli et al. (2012), Gennaioli et al. (2015), and Jin (2015)) and the representativeness heuristic (Gennaioli and Shleifer, 2018). Other sources of tail risk neglect are undersampling extreme events (Hertwig et al., 2004) and assuming a unimodal functional form (e.g., a normal distribution) when learning about fat-tailed payoffs under model uncertainty.\textsuperscript{14} Here we propose an additional source of tail risk neglect which is not learning-related but relates to the way outliers are misperceived in the first instance, even before the learning process can start. By shedding light on a conservative bias

\textsuperscript{12}For example, Squires et al. (1975), Tse et al. (2004), Ferrari et al. (2010).

\textsuperscript{13}For example, Lo (2017) describes in detail in Chapter 9 how the news of Lehman collapse did not come out of nowhere inasmuch as the regime shift that started the global financial crisis happened as early as in June 2006, when home prices peaked and then started to fall. Rising interest rates and decreasing prices led to defaults in homeowners, and the ensuing stop in the flow of money that supported the pools of MBS securities was a disaster leading the “dominos” of the economy to fall one by one, starting by the bankruptcy of risk-taking mortgage lenders.

\textsuperscript{14}For example, Taleb (2004b), Donnelly and Embrechts (2010), and Payzan-LeNestour (2018).
whereby outliers are perceived as less extreme than in reality, the present work may help in better understanding the mechanisms underlying financial crises. For example, outlier blindness hypothesis provides a natural explanation for the complete lack of reaction of investors to the regime shift news of the housing bubble burst during the pre-Lehman period, which is puzzling at first glance. During that period, which has been referred to as “the quiet period” (Gennaioli and Shleifer, 2018), the evidence for a major regime shift was compelling, yet investors continued to neglect tail risk (including when daily realized volatility started to rise above 20% from the summer of 2007, with peaks above 40%). One explanation for this apparent paradox is that investors chose to ignore the 2006 regime shift news for greediness reasons (Lo, 2017). The present findings point to a complementary explanation, which is that investors grossly underestimated the importance of the news—consistent with the foregoing conservative bias—as a consequence of not being yet adapted to the new regime.

I. Theory

A. Efficient coding principle

FROM OBJECTIVE TO PERCEIVED STIMULUS MAGNITUDES

A common way of modeling human perception is to suppose that when a stimulus with true magnitude \( x \) (on some dimension of interest) is presented, an internal representation is produced, the information in which can be summarized by a sufficient statistic \( r \), which we assume for simplicity to also be a single real number. The representation \( r \) is drawn from a probability distribution \( p(r|x) \), conditional on the true stimulus magnitude.

If two stimuli \( x_1 \) and \( x_2 \) are presented and the subject must decide which magnitude is greater, subject judgment is based on the internal representations \( r_1 \) and \( r_2 \) evoked by the respective stimuli. The subject responds that the second stimulus is greater if and only if \( r_2 > r_1 \).

\(^{15}\text{See Lo (2017) and Gennaioli and Shleifer (2018).}\)
\(^{16}\text{In a more realistic account, the internal representation will be a high-dimensional object, describing a complete pattern of neural activation; but we assume here that a single sufficient statistic } r \text{ describes the evidence in this pattern that is relevant for discriminating between two stimuli along the dimension of interest.}\)
\(^{17}\text{This is the optimal decision rule to maximize the frequency of correct responses if the conditional probabilities satisfy the monotone likelihood ratio property (i.e., for any true magnitudes } x, x' \text{ such that } x' > x, \text{ the likelihood ratio } p(r|x')/p(r|x) \text{ is an increasing function of } r \text{) and if any pair of stimuli are equally likely to presented in either order (i.e., under the prior probability distribution for the true values } (x_1, x_2), (x, x') \text{ and } (x', x) \text{ are equally likely, for any magnitudes } x, x'.\)
The probability of an erroneous judgment 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 Figure 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).

Importantly, the mean of the conditional distribution is not the objective stimulus value $x$ but a continuously increasing function of $x$, which we denote by $m(x)$. To reflect the fact that the degree of precision with which different stimuli can be distinguished is limited by the available resources of the brain (we have a finite number of neurons and a finite number of possible spike outputs of each neuron), the function $m(x)$ takes values within some bounded interval $[\underline{m}, \overline{m}]$.

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 stimulus space requires a corresponding reduction in the
number that can be effectively discriminated elsewhere; therefore, it is efficient to allocate fewer possible distinctions to regions of the stimulus space that are infrequently encountered, given that the capacity to make fine distinctions in the latter case will seldom matter. Put it differently, an efficient use of finite resources requires finer distinctions to be made over those parts of the stimulus space where stimuli are most frequently encountered. In natural environments, the frequency distribution of a particular magnitude is a unimodal function, such as a Gaussian distribution. This means that the \( m(x) \) function has a sigmoid shape, increasing most steeply over an intermediate range near the mode of the frequency distribution. For example, the \( m(x) \) function in Figure 2 (bottom graph) represents an efficient coding scheme in the case of an environment in which the distribution of stimulus values is given by the density function \( f(x) \) (Figure 2, top graph). See how \( m(x) \) 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 (Figure 2, bottom graph) represents an efficient coding scheme when the distribution of stimulus values is given by \( \tilde{f}(x) \) (Figure 2, top graph): \( \tilde{m}(x) \) increases most steeply at the value of \( x \) where \( \tilde{f}(x) \) is highest, and progressively less steeply for more extreme values.

Outlier blindness

It is clear from the example shown in Figure 2 that the slope of the \( m(x) \) (resp. \( \tilde{m}(x) \)) function is null in the tails of the frequency distribution \( f(x) \) (resp. \( \tilde{f}(x) \)), which means that values of \( x \) that are extreme relative to the prior (whether values in the far left tail or the far right tail) are poorly distinguished from one another. Figure 2 also illustrates the fact that what counts as an outlier depends on what one has reason to expect. The same two stimuli \( x_1 \) and \( x_2 \) are not outliers in the environment in which the frequency distribution of stimuli is described by \( f(x) \) and hence they are are fairly accurately discriminated, whereas they 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. The core of the experimental test described in Section II exploits this feature of efficient coding.

\[18\] In Appendix V.A, we present an explicit constrained optimization problem, with a performance measure appropriate to the task in our experiment, that allows such a conclusion to be obtained. For examples of similar conclusions in sensory contexts, based on alternative performance measures, see Laughlin (1981), Wei and Stocker (2015), and Ganguli and Simoncelli (2016). For example, Laughlin (1981), Robson (2001), Woodford (2012), Polania et al. (2019).
Figure 2. Two possible distributions for the stimulus magnitude $x$ ($f(x)$ and $\tilde{f}(x)$, top graph), and their corresponding efficient coding schemes (respectively $m(x)$ and $\tilde{m}(x)$ on bottom graph). Functions $m(x)$ and $\tilde{m}(x)$ have the same range of possible stimuli (the horizontal axis $x$ in the figure), and the same bounded range of variation. Nonetheless, the degree to which the two magnitudes $x_1$ and $x_2$ are discriminated is different in the two cases. It is easier to tell stimuli 1 and 2 apart in the case of function $m(x)$ ($\mu_2 - \mu_1$ large) relative to the case of function $\tilde{m}(x)$ ($\tilde{\mu}_2 - \tilde{\mu}_1$ small, implying the two stimuli are frequently confused).
Outlier blindness leads to under-estimating the probability of tail events. To appreciate this point, note that a subjective estimate $\hat{x}$ of a stimulus magnitude $x$ is necessarily a function of the internal representation $r$, but $\hat{x}(r)$ need not be a linear function of $r$. Actually, under the hypothesis that the subject’s responses are optimally adapted to the statistics of her environment (a hypothesis of optimal decoding in addition to efficient encoding), the estimation function $\hat{x}(r)$ should generally be a nonlinear transformation of $r$, and one that at least to some extent “undoes” the effects of the $m(x)$ function. In the absence of any encoding noise (i.e., the standard deviation of $p(r|x)$ is null), the optimal decoding rule would be simply $\hat{x}(r) = m^{-1}(r)$, so that $\hat{x} = x$ for all $x$ (even extreme outliers).

However, in the case of finite-precision encoding, even with optimal decoding the perceived magnitude of outliers will generally be less extreme than the reality. This is illustrated by the numerical examples presented in Woodford (2012), Polania et al. (2019), and Frydman and Jin (2018), who find that when the decoding rule $\hat{x}(r)$ is computed to minimize the mean squared error of the estimates, the estimates involve $\hat{x} < x$ on average for all large enough values of $x$, and $\hat{x} > x$ for all small enough values of $x$ (“conservative bias”). Hence if the available memories of past values upon which perceptions of tail risk are based are simply estimates based on noisy perceptions of the kind discussed here, decision makers should be expected to under-estimate the frequency of occurrence of extreme values in either direction.

Note

There are two important notes about the efficient coding scheme described in Figure 2. First, it relies on adaptation, i.e., its implementation requires that the brain learns the density associated with the current environment ($f(x)$ or $\bar{f}(x)$); such learning works through repeated exposure (“adaptation”) to stimulus values drawn from this density. Second, this efficient coding scheme is only optimal with unimodal densities. It is not meant for environments featuring multimodal densities.

---

20 Examples of models of magnitude estimation that incorporate both efficient coding and optimal (Bayesian) decoding include Woodford (2012), Wei and Stocker (2015), Polania et al. (2019), and Frydman and Jin (2018).
B. Does outlier blindness occur in unstable settings? “Outlier Blindness Hypothesis’

As stressed in the Introduction, the occurrence of outlier blindness in modern economic environments is not a foregone conclusion given that payoff distributions are multimodal, frequently shifting over time. The foregoing note points to two key conditions for outlier blindness to occur in such environments: i) adaptation occurs quickly in humans; ii) the previously described efficient coding scheme is implemented irrespective of the statistics of the environment, i.e., even under multimodal stimulus distributions. In light of the latest literature on adaptation and efficient coding, we make the following two assumptions which imply that both conditions are met.

Assumption i): Range adaptation

How much experience with the frequency distribution \( f(x) \) associated with a given environment is required for the brain to be able to implement the kind of efficient coding scheme described above? There is ample evidence that neural coding appears to be recalibrated quickly to a new environment, suggesting adaptation does not require much experience.\(^{21}\) This seems counterintuitive if on assumes that recalibration requires re-learning the entire frequency distribution \( f(x) \) each time it changes. However, the evidence for quick recalibration is consistent with one specific form of efficient coding in which the previously described efficient coding scheme is optimized within a two-parameter family of possible functions \( m(x) \). Specifically, we restrict the set of possible efficient coding schemes to the ones of the form \( m(x) = \tilde{m}(\phi(x)) \), where

\[
\phi(x) = \alpha + \beta x
\]

is an affine transformation of the stimulus space \( \beta > 0 \), and \( \tilde{m}(x) \) is a sigmoid function of the kind shown by \( m(x) \) or \( \tilde{m}(x) \) in Figure 2. Then \( \alpha \) and \( \beta \) parameterize a two-parameter family of possible encoding rules, in all of which \( m(x) \) lies within the bounds \([\tilde{m}_1, \tilde{m}_2]\) for all \( x \). The hypothesis of optimal range adaptation proposes that for any distribution of environmental values \( f(x) \), the encoding rule is given by the particular rule within a two-parameter family of this kind that maximizes expected performance in that environment.\(^{22}\) Under this rule, the tuning slope of neurons’ firing rate to encode stimulus

\(^{21}\)See the references in Introduction and below.

\(^{22}\)For examples of range-adaptation models of this kind, see Soltani et al. (2012) and Rustichini et al. (2017).
value is “range-adapted,” i.e., it is inversely proportional to the range of stimulus values available in the current environment, thereby ensuring that the firing rates of neurons span the full activity range in each environment (see Figure 3). The fact that the full activity range is always available to represent the range of stimulus values offered in the current environment is one key aspect by which range adaptation is “efficient.” Importantly, range adaptation is observed for all kinds of stimuli including economic value (e.g., Tobler et al. (2005), Padoa-Schioppa and Rustichini (2014), Rustichini et al. (2017)).

A key characteristic of the range adaptation model is that only two parameters of the frequency distribution associated with a given environment, namely, a location parameter and a measure of dispersion (for example, the mean and standard deviation; or alternatively, the two boundaries of the inter-quartile range), are to be updated in order for the perceptual system to adjust to changes in the statistics of the environment. This means that even a small sample of observations from a new environment might well be sufficient for adaptation to occur. There is a good deal of evidence for neural coding that adapts to the statistics of the environment in this way. The best-known examples involve quick shifts in neural coding in response to changes in the mean stimulus value, as well as

---

23 A classical reference is Laughlin and Hardie (1978) which finds that the stimulus intensity required to produce a given physiological response increases with increases in the average stimulus intensity.
quick adaptation to changes in the variability of stimulus value. Therefore, having the range adaptation model of efficient coding in mind, one expects adaptation to occur even when the frequency distribution $f(x)$ shifts very frequently over time as modelled in the experiment documented in Section II.

Assumption ii): Unimodal priors

We assume that the brain uses unimodal prior distributions for stimulus magnitude even when the true distribution is multimodal—so using a multimodal prior would be optimal to match the statistics of the environment. The idea here is that the human brain evolved to maximise survival chances in the kind of statistical environments in which it emerged, namely, natural environments. The distributions for the magnitude of most stimuli encountered there are unimodal. Viewed from evolutionary timescales, the change to multimodal distributions with the appearance of modern economic environments occurred yesterday. This makes it unlikely that the human brain was reconfigured to this change. So one should expect the brain to stick to unimodal priors irrespective of the statistics of the current environment.

If Assumptions i) & ii) are true, this suggests that outlier blindness occurs in modern economic environments despite the instability. The lab experiment reported in Section II is a direct test of this hypothesis.

II. Lab experiment

A. Experimental Design

Task Description

The principle of the task is that on each of 1,128 trials, the subject is to discriminate between the shades of grey of two rectangles by clicking within 2 seconds on the rectangle that looks darker, or on an “=” icon displayed in the middle of the screen if the rectangles look of the same color (see Figure 4). We construct a scale of grey with 12 different shades (1: very light; 12: extremely dark). In a typical trial, the shades of the two rectangles are adjacent, i.e., shade values $x$ and $x + 1$ are displayed, where $x$ is randomly drawn.

\(^{24}\) A classical reference is Fairhall et al. (2001) which provides evidence that the size of difference between two stimulus magnitudes required to produce a given difference in physiological response scales with the standard deviation of stimulus magnitudes in a given environment. Additional references are provided in Soltani et al. (2012).
from the scale of grey as described next. The two shades are the same in 13% of the trials. The subjects are told the frequency of same-shade trials in the task instructions (see Appendix V.B).

The logic of our design is to have each subject go through a 40-trial “adaptive sequence,” in which on each trial shade value \( x \) is drawn from a normal distribution with mean \( m \) and standard deviation \( s \) (the values of \( m \) and \( s \) are randomly selected in the sets \{3, 4, ..., 10\} and \{1, 2\} respectively), immediately followed by an “adaptive test trial” in which the shade \( (x') \) is randomly drawn in the range of values located at least three standard deviations from \( m \). So by design, the shade presented at the adaptive test trial is an outlier as it falls outside the range of values to which the subject has been adapted in the adaptive sequence. Each outlier value \( x' \) defines a 40-trial “control sequence” in which the shade value on each trial is normally distributed around \( x' \) (standard deviation is either 1 or 2; choice is random). The control sequence systematically ends with a “control test trial” in which \( x' \) is presented. So each adaptive sequence is paired with a control sequence and both sequences are immediately followed by the same test trial.

There are 12 adaptive sequences (so 12 sequence pairs). The order of appearance of all sequences is randomized. So generally, the adaptive and control sequences of a given pair do not follow each other, and we do not impose either that the adaptive sequence be presented before the control sequence.\(^{26}\)

For each subject, we compute the mean accuracy level in the control test trials averaged across the 12 sequence pairs, as well as the mean accuracy level in the adaptive test trials. Accuracy is defined as the fraction of correct replies. Our main statistic of interest is the difference between the accuracy level in the control vs. adaptive test trials. To double the number of observations for our main statistic without increasing the total number of sequence pairs, we add to our design a symmetrical feature whereby within each sequence pair, the shade value presented at the last trial of the adaptive sequence is also presented immediately after the control test trial. When presented within the adaptive sequence, the value is not an outlier whereas when presented after the control test trial, it is an outlier as by design it falls at least three standard deviations away from the mean value to which the subject has been adapted during the control sequence. The last trial of the adaptive sequence thus serves as a control test trial w.r.t the adaptive test trial that is

\(^{25}\)The drawn value is rounded. For instance, if the number drawn is 4.3, the 4th-shade on our scale of grey is chosen.

\(^{26}\)The only constraints that we impose are that the means of two successive sequences be as close as possible (to ensure the absence of outliers at the beginning of a sequence) and that each \( m \) value from the set of possible values \{3, 4, ..., 10\} be used in at least two of the adaptive sequences.
put immediately after the control test trial. So for each subject, we have 24 observations for our main statistic of interest.

Outlier blindness hypothesis predicts a significantly decreased accuracy in the adaptive test trials compared to the control test trials. The differential accuracy comes from the fact that by design, the shade value presented at the adaptive test trial is improbable from the perspective of the subject inasmuch as it is well outside the range of values the subject is expecting, whereas when presented in the control test trial, the same shade value is probable from the perspective of the subject (it is within the range of values the subject is expecting). As explained in the introduction, efficient coding theory predicts that perception accuracy is decreased for improbable values (e.g., Tobler et al. (2005) and Woodford (2012)).

Of note, the proportion of same-shade trials is the same (13%, as noted above) across all trial types (the trials within the adaptive sequences, the trials within the control sequences, and the test trials). This is important to allow comparisons across trial types, as one expects the base rate accuracy to be decreased in the same-shade trials (irrespective of trial type) inasmuch as the subjects know that same-color trials occur with low probability in this task, so in principle their replies should be biased against the “=” reply.

**WHY THE CURRENT TASK SPECIFICATION**

As stressed in the Introduction, our motivation for designing a non-economic task in order to study the perception of financial returns in economic agents is motivated by a growing consensus in neuroscience that efficient coding applies similarly to the perception of simple sensory stimuli and the perception of economic value. This suggests that one can apply findings from the former context to the latter. By using a simple perceptual task rather than an economic decision-making task, we avoid potential confounds related to risk attitude which could not be easily controlled for in an economic decision-making task (fully disentangling risk preferences and beliefs is not possible).

Our main statistic of interest allows us to isolate the outlier blindness effect by ruling out any confound related to the extreme-value effect as explained in the Introduction (the fact that outliers are more likely to be located at the extremes of the scale, and it is harder to discriminate between shade values located at the extremes). If accuracy is inherently decreased when discriminating between extreme values, this affects both adaptive and control test trials in the exact same way in this task, since adaptive and control test trials
Our main motivation for choosing the current task settings is to maximize statistical power in the analysis. For instance, our choice of 2 seconds as the allowed time to make a choice on each trial is to maximize the number of trials for each subject under the standard duration limits for this kind of task (30 minutes should not be exceeded), while keeping in mind task feasibility (excessive time pressure leads to random choice in the subjects). We tested different time parameter values in pilot sessions; both subject choice data and subject oral reports on the task after the pilot suggest that allowing 2 seconds to make a choice is a reasonable trade-off between sample size and task feasibility. Subject behavior in the experiment also validates this parameter choice inasmuch as subjects usually reply well within the allowed time, the frequency of missed trials is fairly low overall, and doubling the allowed time to make a choice does not change our main findings (more in the Results section). All this suggests time pressure is not an issue in this experiment.

Likewise, our motivation for designing a 3-choice discrimination task (the default would be a 2-choice discrimination task with no same-color trials) is to maximize statistical power by averting some “ceiling effect” in accuracy level which would emerge if the task

---

**Figure 4. User interface of the experimental task.** The task is a 3-choice discrimination task in which on each trial the subject is to click on the rectangle that looks darker, or on an “=” icon displayed in the middle of the screen if the rectangles look of the same color. The time allowed to make a reply is 2 seconds (remaining time is indicated through a timer at the bottom of the screen).
were too easy. Our choice of a 12-point scale for our scale of grey follows similar logic. Using a coarser scale would potentially decrease our power by making the task too easy. Using a finer scale would increase randomness in choice at the test trials (the finer the scale, the harder it is to discriminate between extreme values). We tried different degrees of coarseness in pilot sessions and arrived at the present scale. 

Last but not least, one important aspect of reducing noise in the data is to ensure that the subjects do pay attention on each trial. To that purpose, we use a special experimental procedure—spelled out next.

### B. Experimental procedure

Sixty-nine undergraduate students from the University of New South Wales register online to participate in the experiment. Upon arrival at the lab, they watch the online instructions for the experiment for 15 minutes (see Appendix V.B). At the very end of the instructions, they complete a 3-minute training session in which they play a few trials of the experimental task to familiarize themselves with the task interface (more on this below), after which they are briefed again by the experimenter on the distinctive nature of the experiment, the payment rule that is used in particular. The experimenter re-emphasizes that each correct reply yields $0.1, each incorrect reply leads to a loss of $0.25, each missed trial leads to a loss of $1, and that the subject receives at the end of the experimental session all the net accumulated outcomes from the task. The experimenter also stresses that the luminance setting has been adjusted on each machine before the beginning of the experimental session and that it has been locked so that luminance cannot be changed during the experiment. Subsequently, the subjects complete one run of the task, which lasts for approximately 25 minutes.

To minimize noise in our data by ensuring that the subjects do pay attention on each trial, we take two steps. First we provide our subjects with high monetary incentives by rewarding high performance in the task through high payoffs. As indicated above, each correct reply yields $0.1 and the subject performs 1,128 trials overall, so high performers

---

27 We ran a pilot session of the 2-choice variant of the task with no same-color trials and found the overall accuracy was very high.
28 We do not claim that our choice is optimal but our findings show it is good enough to identify the outlier blindness effect which is our main focus in this study.
29 Registration is done through the ORSEE recruitment system (http://www.orsee.org/web/).
30 In case of negative earnings, the subject ends up with the $5 show-up reward which is given irrespective of task performance (as per the lab protocol).
31 The monitors in the lab are the HP E272q. We used the factory default settings (brightness: 90; contrast: 80; dynamic contrast: off; black stretch: off). To prevent participants from changing settings during the experiment, the experimenter enabled the button lockout feature of the monitors beforehand.
can potentially end up with more than $100. This feature of the experiment is well emphasized in the task instructions. Eight subjects earn more than $70, 32 earn more than $50 (mean: 43.8; median: 50.10; mode: 45.65; std: 36.7; see Appendix V.C). In addition to rewarding high performance in the task through high payoffs, an important aspect of incentivizing the subjects consists of using a “pay all” payment rule (we pay for the outcome from every decision made) rather than the alternative “pay one” approach, which pays for the outcome of only a subset of the choices made. The pay all approach averts the issue of diluted incentives which arises if, for instance, the subjects are susceptible to the “disjunction effect” or the like.\textsuperscript{32}

The second key feature of the experimental procedure used in this experiment consists of using the online instructions not only in the usual way (i.e., to acquaint the subjects with the experimental task) but also as a screening device in order to screen out subjects that are not suitable for the task. Specifically, at the very end of the online instructions, the subjects are to decide whether they wish to do the aforementioned short training session to get familiar with the task interface before proceeding to the test. The decision to skip the training, which is implemented by clicking on a button “I wish to skip the training session and go directly to the task” (see Appendix V.B), results in being immediately excluded from the experiment.\textsuperscript{33} The logic here is that the decision of a given subject to skip the training session provides a reliable signal that the subject is not highly motivated to do well in the task. Six subjects choose to skip the training and are therefore excluded from the experiment.

\section{III. Results}

\textbf{A. Findings of the Original Experiment}

Across subjects, we find that accuracy is significantly lower in the adaptive test trials than in the control test trials, as predicted by outlier blindness hypothesis. Our statistic of interest (the accuracy difference between control and adaptive test trials, as explained above) averaged across subjects is significantly positive according to our statistical tests (e.g., paired t-test based on the individual statistics of interest as defined above: $t = 11.7; p \sim 0$, two-tailed); see Figure 5 and Figure 14 in Appendix V.D. Strikingly, the statistic is positive for almost all (60 out of 63) the subjects; see Figure 6.

\textsuperscript{32}For example, Tversky and Shafir (1992), Shafir and Tversky (1992), Shafir (1994).

\textsuperscript{33}The experimenter gives the excluded subjects the $5 show-up reward.
The mean accuracy level across all subjects and trial types is 0.83. It is markedly decreased in the same-shade trials (0.44), which as explained above is expected given the nature of the task. Accuracy is significantly higher on average within the adaptive sequences than within the control sequences (0.87 vs. 0.81), which comes as no surprise in light of the aforementioned extreme-value effect.\(^{34}\) The average response time in the experiment is 0.9 sec (min: 0.66; max: 1.41; std: 0.14). There is a negative correlation between accuracy and response time (one way ANOVA test: \(f = 1454; p < .00001,\) two-sided). The percentage of missed trials is very low on average (0.005\%). Appendix V.C provides more details on the main descriptive statistics related to missed trials, response times, subject accuracy, and subject earnings. Appendix V.D provides more details on the main statistical tests.

B. Follow-Up Experiments

Placebo Experiment

An obvious prerequisite for the outlier blindness effect is that the agent holds expectations. An “outlier” is indeed defined with respect to the expectation held by the agent. If the agent expects nothing in particular, the notion is meaningless (there are no outliers from the perspective of the agent). According to efficient coding theory, expectations are set through adaptation. Applied to the current experiment, what this means is that following each 40-trial sequence, the subjects expect the range of shade values to which they have been exposed during the sequence. Therefore, a litmus test of outlier blindness hypothesis consists of suppressing adaptation in a follow-up “placebo experiment” in which the length of each sequence is reduced to a very low level, so that the subjects do not have time to form expectations about shade values, and hence there should be no outlier blindness effect.

In the placebo experiment (N = 35; same cohort as in the original experiment: undergraduate students at the University of New South Wales), each sequence lasts for only 3 trials. There are 160 sequences (960 trials overall), so 320 observations for our statistic of interest. Except for sequence duration (and total number of sequences), the experimental task is the same as the one used in the original experiment. Appendix V.C reports the

\(^{34}\)As explained above, the extreme-value effect refers to the fact that it is harder to discriminate between shades located at the extremes of the scale. By design, the shades that are presented within the control sequences are more extreme on average than those presented within the adaptive sequences. So the extreme-value effect implies that accuracy be decreased on average within the control sequences. See Figure 16 in Appendix V.D for more details.
Figure 5. Accuracy in adaptive test trials (left) and control test trials (right). Heights of bars indicate the mean accuracy averaged across the 63 subjects. Accuracy is defined as the fraction of correct replies. Line segments indicate SEM. **** $p < .00001$. 
Figure 6. Comparative accuracy levels in adaptive versus control test trials. Each data point corresponds to one subject (N=63; 24 observations per subject). x axis: accuracy in the adaptive test trials. y axis: accuracy in the control test trials. Accuracy is defined as the fraction of correct replies. Data points above the 45 degree line correspond to subjects for whom accuracy is decreased in the adaptive test trials, as predicted by outlier blindness hypothesis.
main descriptive statistics with regard to response time, earnings, and accuracy per trial type in the placebo experiment.

Figure 7 displays the distribution of accuracy in the adaptive and control test trials across subjects, for both the original and placebo experiments. The main effect documented above for the original experiment is apparent on this graph (blue curves: the mean accuracy level at the adaptive test trials is shifted to the left relative to the mean accuracy at the control test trials). In contrast, the effect is absent in the placebo experiment (red curves: the distributions of adaptive and control test trials accuracy overlap). One cannot reject the null hypothesis that in the placebo experiment, the mean accuracy level is the same in the adaptive and control test trials, as predicted by outlier blindness hypothesis (see Figure 14 in Appendix V.D).

Figure 7. Density plot of accuracy in the adaptive test trials and control test trials, for both the original experiment (blue) and the placebo experiment (red). The density is derived across subjects, based on 63 subjects, 24 observations per subject (original experiment), and 35 subjects, 320 observations per subject (placebo experiment).

Note that variance is lower for the placebo experiment than for the original experiment (the density width is lower for the placebo experiment, for each trial type). This is because sample size is much higher, namely 320 observations per subject vs. only 24 in the original experiment (no time to do more to ensure task duration does not exceed 30 minutes).
SHORTENED ADAPTATION

Our finding that the outlier blindness effect is suppressed in the placebo experiment rules out the possibility that our statistic of interest in the original experiment be driven by some hidden experimental artefact (since the task used in both experiments is the same except for the adaptation length factor, as explained above). As such, it strengthens the evidence for outlier blindness hypothesis.

Next we ask how long it takes for the human brain to hold expectations. Following Assumption i) proposed in Section I, we hypothesize that increasing adaptation length by a few trials is enough to restore the outlier blindness effect. To test this hypothesis, we run the “5-trial adaptation experiment” (N=31; same cohort as in the original experiment: undergraduate students at the University of New South Wales) in which each sequence lasts for 5 trials. There are 100 sequences in that experiment (1,000 trials overall).36 Consistent with our hypothesis, we find the outlier blindness effect is significant in the 5-trial adaptation experiment. The null hypothesis that accuracy is the same in the adaptive and control test trials is rejected in each of our statistical tests (e.g., paired t-test: \( t = 4.2, p < .001 \), two-tailed; see Figure 14 in Appendix V.D).

This finding shows that an adaptation length of five trials is sufficient for the agent to form expectations and hence for the outlier blindness effect to reappear. Note that the original effect is only partially recovered: the outlier blindness effect is four times stronger in the original experiment than in the 5-trial adaptation experiment (two sample t test to compare the mean statistic of interest in the two experiments: \( t = 8.8, p < .001 \); see Figure 15 in Appendix V.D). This again conforms to the theory proposed in Section I which predicts that the magnitude of the outlier blindness effect increases with adaptation length; adaptation length is eight times bigger in the original experiment than in the 5-trial adaptation experiment.

ROBUSTNESS CHECKS

We find that the foregoing findings are robust to excluding the data from the beginning of the task, to account for the possibility that subjects need some time to become familiar with the task interface.37 The main accuracy results are also unchanged when we exclude

---

36See Appendix V.C for the main descriptive statistics regarding response time, earnings, and accuracy per trial type in the 5-trial adaptation experiment.

37Note that a priori the subjects are familiar with the task interface from the start, since all of them complete a 3-minute training session prior to the task as per the experimental procedure described in Section II.B.
Outlier Blindness

In the final stage of the study, we examine whether the outlier blindness effect is affected by time pressure. We run a final experiment (N= 33; same cohort as in the original experiment: undergraduate students at the University of New South Wales) in which the time allowed to provide a reply on each trial is doubled relative to the original experiment (4 seconds vs. 2 seconds). The average response time in the experiment with double response time is 1.20 seconds on average across subjects, which is significantly above the mean response time in the original experiment (two sample t-test: $p \sim 0$). However the magnitude of the outlier blindness effect is unchanged in the experiment with double response time compared to that in the original experiment. This is already apparent from Figure 8 which displays the distribution of accuracy in the adaptive and control test trials across subjects, for both the original experiment and the experiment with double response time. The distributions for each trial type look very similar across the two experiments. Formal statistical tests confirm that doubling the response time allowed on each trial does not affect the outlier blindness effect, which remains highly significant (see Figure 14 in Appendix V.D). We cannot reject the null hypothesis that the outlier blindness effect is similar in the two experiments (two sample t-test to compare the mean statistic of interest in the two experiments: $t = 1.9; p = 0.056$; see Figure 15 in Appendix V.D).

IV. Conclusion

In this study, we hypothesize that economic agents are hampered in their perception of extreme values because the human brain evolved to make fine distinctions between values in the expected range and only coarse distinctions (if any at all) between outliers. This phenomenon, which we call “outlier blindness,” is maladaptive in unstable settings such as modern financial markets, where extreme events are not rare enough for their neglect to be innocuous for economic agents (Taleb, 2004a). But it was adaptive (“efficient”) in the ancestral human lifestyle, where the distribution of payoffs was most often unimodal. As such, outlier blindness hypothesis is reminiscent of the idea that our brains evolved to optimize survival chances in the environmental context in which they emerged. Taken out

38See Appendix V.C for the main descriptive statistics regarding response time, earnings, and accuracy per trial type in the experiment with double response time.
39An exception is the case of outliers that should be ignored altogether as they are pure aberrations caused by leptokurtic noise (Bossaerts and d’Acremont, 2016). The 1987 stock market crash is one example of such “aberrations.” Traders overreacted to it, thinking that something fundamental had changed, which was not the case (within 1 year, the stock market had more than recovered).
Figure 8. Density plot of accuracy in the adaptive test trials and control test trials, for the original experiment (blue) and the experiment with double response time (orange). The density is derived across subjects, based on 63 subjects, 24 observations per subject (original experiment), and 33 subjects, 24 observations per subject (placebo experiment).
of their proper environment, they become maladaptive (see Robson (2001) and Chapter 6 of Lo (2017)).

To test outlier blindness hypothesis, we use a special experimental paradigm that allows us to separate the extreme-value effect (which simply reflects the fact that it is harder to see at the extremes on a scale of values, as described by the well-known Weber-Fechner law) from the outlier blindness effect which is the novel focus in this study. Our design further allows us to disentangle the outlier blindness effect from other potential factors related to risk preferences.

We provide strong evidence in favor of outlier blindness hypothesis. We find that the magnitude of the bias is highly significant in unstable environments (when payoff distributions shift every 40 trials) and still significant in very unstable environments (when payoff distributions shift every 5 trials). In regard to human psychology, these findings suggest that humans are quick to form expectations. In regard to practical implications for financial agents, they suggest that the outlier blindness phenomenon is pervasive, prevailing over all kinds of environments in terms of instability levels, meaning that no one should expect to be immune to the outlier blindness effect, including those operating in extremely unstable environments.
V. Appendix

A. Efficient Coding

To come soon.

B. Instructions for the Task

The reader will find below the text of the instructions provided to the subjects in the original experiment. The text of the instructions for the placebo and 5-trial adaptation experiments is the same except for the information concerning the total number of trials (there are 960 trials in the placebo experiment and 1,000 trials in the 5-trial adaptation experiment). The text of the instructions for the experiment with double response time is the same except for the information about the time allowed to make a choice on each trial (which is 4 seconds in the double-response time experiment).
Thanks for accepting to participate in “The Hue Task”!

On each of 1,128 trials, you will be asked to discriminate between two hues of grey, see picture below. In some of the trials the two rectangles will be of the same color. In some others, they will be of different color. If the two rectangles look of the same color, click on the “=” icon in the middle of the screen. If the rectangles look of different color, click on the rectangle that looks darker.

Which one is darker?

We (the experimenters) want to reward very significantly high performance in this task. You will earn $0.10 per correct reply and will lose $0.25 per incorrect reply. Since you will be playing 1,128 trials overall, this means you can earn a lot of money potentially—more than $100—if you perform well in the task.

Please Note:
- You will lose $1 if you fail to reply within the impared time on a given trial (2 sec — note the pace of the game is high).
- The task is quite long (about 30 minutes overall). There will be a short break after the first 15 minutes.

The task therefore requires you to keep the pace and pay attention for a prolonged period of time. To familiarize you with the task settings before performing the task, we offer you the opportunity to do a 3-minute training session in which you will be playing a few trials of the task; note these trials will NOT be counting for your final payment (your replies won’t be recorded).

Please indicate your choice:

- I wish to do the training session before performing the task
- I wish to skip the training session and go directly to the task

Figure 9. Instructions for the task.
C. Descriptive Statistics for Each Experiment

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Missed Trials</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment</td>
<td></td>
<td>6.17</td>
<td>10.04</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>5-Trial Adaption Experiment</td>
<td></td>
<td>7.52</td>
<td>10.32</td>
<td>0</td>
<td>47</td>
</tr>
<tr>
<td>Placebo Experiment</td>
<td></td>
<td>9.29</td>
<td>5.82</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>Experiment with Double Response Time</td>
<td></td>
<td>1.52</td>
<td>1.89</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 10. Number of missed trials in each experiment. The mean, standard deviation, min and max are derived across subjects, for each experiment.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mean Accuracy</th>
<th>Overall</th>
<th>Same Color Trials</th>
<th>Different Color Trials</th>
<th>Adaptive Trials</th>
<th>Control Trials</th>
<th>Adaptive Test Trials</th>
<th>Control Test Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment</td>
<td>0.83</td>
<td>(0.01)</td>
<td>0.44</td>
<td>0.88</td>
<td>0.81</td>
<td>0.61</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>5-Trial Adaption Experiment</td>
<td>0.80</td>
<td>(0.01)</td>
<td>0.41</td>
<td>0.83</td>
<td>0.84</td>
<td>0.72</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Placebo Experiment</td>
<td>0.77</td>
<td>(0.01)</td>
<td>0.29</td>
<td>0.79</td>
<td>0.86</td>
<td>0.85</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>Experiment with Double Response Time</td>
<td>0.90</td>
<td>(0.01)</td>
<td>0.43</td>
<td>0.95</td>
<td>0.92</td>
<td>0.89</td>
<td>0.72</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Figure 11. Mean accuracy according to trial type. “Adaptive trials”: trials within the adaptive sequences. “Control trials”: trials within the control sequences. Numbers in parenthesis: sem.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Response Time (sec)</th>
<th>Overall</th>
<th>Same Color Trials</th>
<th>Different Color Trials</th>
<th>Adaptive Control Trials</th>
<th>Adaptive Test Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment</td>
<td>0.90</td>
<td>1.08</td>
<td>0.87</td>
<td>0.86</td>
<td>0.92</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>5-Trial Adaption Experiment</td>
<td>1.03</td>
<td>1.16</td>
<td>1.02</td>
<td>0.99</td>
<td>1.01</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Placebo Experiment</td>
<td>1.09</td>
<td>1.24</td>
<td>1.08</td>
<td>1.03</td>
<td>1.04</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Experiment with Double Response Time</td>
<td>1.20</td>
<td>1.81</td>
<td>1.12</td>
<td>1.15</td>
<td>1.23</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.07)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.05)</td>
</tr>
</tbody>
</table>

**Figure 12.** Mean response time according to trial type. “Adaptive trials”: trials within the adaptive sequences. “Control trials”: trials within the control sequences. Numbers in parenthesis: sem.

**Figure 13.** Earning distribution across subjects in each experiment. For each subject the earnings are computed as the net accumulated outcomes at the end of the task.
### D. Main Statistical Tests

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mean Accuracy in Adaptive Test Trials</th>
<th>Mean Accuracy in Control Test Trials</th>
<th>TWO TAILED PAIRED T TEST</th>
<th>WILCOXON SIGNED RANK TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment (N=63)</td>
<td>0.61 (0.02)</td>
<td>0.79 (0.01)</td>
<td>t = 11.76, p ~ 0</td>
<td>V = 1858.5, p ~ 0</td>
</tr>
<tr>
<td>5-Trial Adaption Experiment (N=31)</td>
<td>0.72 (0.01)</td>
<td>0.75 (0.01)</td>
<td>t = 4.25, V = 330, p &lt; 0.001</td>
<td>p = 0.006</td>
</tr>
<tr>
<td>Placebo Experiment (N=35)</td>
<td>0.73 (0.01)</td>
<td>0.73 (0.01)</td>
<td>t = 1.12, V = 396, p = 0.272</td>
<td>p = 0.131</td>
</tr>
<tr>
<td>Experiment with Double Response Time (N=33)</td>
<td>0.72 (0.02)</td>
<td>0.84 (0.01)</td>
<td>t = 6.27, V = 390, p &lt; 0.0001</td>
<td>p &lt; 0.0001</td>
</tr>
</tbody>
</table>

**Figure 14.** Tests of the outlier blindness effect, for each experiment. The outlier blindness effect is measured by the differential accuracy in the control test trials vs. adaptive test trials as explained in the main text. Numbers in parenthesis: sem.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>(Mean Accuracy in Control Test Trials) – (Mean Accuracy in Adaptive Test Trials)</th>
<th>TWO SAMPLE T TEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment (N=63)</td>
<td>0.18 (0.02)</td>
<td>t = 1.94, p = 0.057</td>
</tr>
<tr>
<td>Experiment with Double Response Time (N=33)</td>
<td>0.13 (0.02)</td>
<td></td>
</tr>
<tr>
<td>Main Experiment (N=63)</td>
<td>0.18 (0.02)</td>
<td>t = 8.89, p ~ 0</td>
</tr>
<tr>
<td>5-Trial Adaption Experiment (N=31)</td>
<td>0.13 (0.02)</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 15.** Comparison of the magnitude of the outlier blindness effect across experiments. Numbers in parenthesis: sem.
### Table: Experiment Mean Accuracy in Adaptive Trials vs Control Trials

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mean Accuracy in Adaptive Trials</th>
<th>Mean Accuracy in Control Trials</th>
<th>Two Tailed Paired T Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment (N=63)</td>
<td>0.87 (0.01)</td>
<td>0.81 (0.01)</td>
<td>t = 8.85, p ~ 0</td>
</tr>
<tr>
<td>5-Trial Adaptation Experiment (N=31)</td>
<td>0.86 (0.01)</td>
<td>0.84 (0.01)</td>
<td>t = 2.75, p = 0.010</td>
</tr>
<tr>
<td>Placebo Experiment (N=35)</td>
<td>0.86 (0.01)</td>
<td>0.85 (0.01)</td>
<td>t = 0.68, p = 0.502</td>
</tr>
<tr>
<td>Experiment with Double Response Time</td>
<td>0.92 (0.01)</td>
<td>0.89 (0.01)</td>
<td>t = 5.16, p &lt; 0.0001</td>
</tr>
</tbody>
</table>

**Figure 16. Measure of the extreme-value effect.** “Adaptive trials”: trials within the adaptive sequences. “Control trials”: trials within the control sequences. The extreme-value effect refers to discrimination being potentially hampered for values that are at the extremes of the scale relative to the other values. By design, the values presented during the control sequences are on average more extreme than the values presented during the adaptive sequences. (The mean shade value for the control sequences is either 3 – 4 or 9 – 10 in the majority of the cases whereas for the adaptive sequences, the mean shade value is evenly distributed across all possible values {3, ..., 10}.) Therefore, one expects the average accuracy to be higher in the adaptive sequences than in the control sequences. Numbers in parenthesis: SEM.
Mean accuracy according to trial type for each experiment. 

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Trials</th>
<th>Missed</th>
<th>Same Color</th>
<th>Different Color</th>
<th>Adaptive</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Experiment</td>
<td>Include</td>
<td></td>
<td>0.83</td>
<td>0.44</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>5-Trial Adaption Experiment</td>
<td>Include</td>
<td></td>
<td>0.80</td>
<td>0.41</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Placebo Experiment</td>
<td>Include</td>
<td></td>
<td>0.77</td>
<td>0.29</td>
<td>0.79</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Experiment with Double Response Time</td>
<td>Exclude</td>
<td></td>
<td>0.90</td>
<td>0.43</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Main Experiment</td>
<td>Exclude</td>
<td></td>
<td>0.84</td>
<td>0.45</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>5-Trial Adaption Experiment</td>
<td>Exclude</td>
<td></td>
<td>0.81</td>
<td>0.42</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Placebo Experiment</td>
<td>Exclude</td>
<td></td>
<td>0.78</td>
<td>0.30</td>
<td>0.80</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Experiment with Double Response Time</td>
<td>Exclude</td>
<td></td>
<td>0.90</td>
<td>0.43</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

Note: Numbers in parenthesis indicate standard error of the mean.
REFERENCES


