Anticipatory Anxiety and Wishful Thinking

By Jan B. Engelmann, Maël Lebreton, Nahuel A. Salem-Garcia, Peter Schwardmann, and Joël J. van der Weele*

December 6, 2023

Abstract

Across five experiments (N=1,714), we test whether people engage in wishful thinking to alleviate anxiety about adverse future outcomes. Participants perform pattern recognition tasks in which some patterns may result in an electric shock or a monetary loss. Diagnostic of wishful thinking, participants are less likely to correctly identify patterns that are associated with a shock or loss. Wishful thinking is more pronounced under more ambiguous signals and only reduced by higher accuracy incentives when participants' cognitive effort reduces ambiguity. Wishful thinking disappears in the domain of monetary gains, indicating that negative emotions are important drivers of the phenomenon.

JEL classification: C91, D83

Keywords: confidence, beliefs, anticipatory utility, anxiety, motivated cognition

*Engelmann: University of Amsterdam, Tinbergen Institute (email: j.b.engelmann@uva.nl); Lebreton: Paris School of Economics (email: mael.lebreton@googlemail.com); Salem-Garcia: University of Geneva (email: salemnahuel@gmail.com); Schwardmann: Carnegie Mellon University (email: schwardmann@cmu.edu); van der Weele: University of Amsterdam, Tinbergen Institute (email: j.vanderweele@uva.nl). We thank the editors, four anonymous referees, Douglas Bernheim, Andrew Caplin, Mark Dean, Yves Le Yaouanq, George Loewenstein, Nathaniel Neligh, Matthew Rabin, Klaus Schmidt, Simeon Schudy, Claudia Senik, Severine Toussaert, and Roberto Weber for useful comments. Li-Ang Chang, Sára Khayouti and Nik Mautner Markhof provided excellent research assistance. We received financial support from the Bavarian Academy of Sciences and Humanities, the DFG's CRC TR190, the Dutch Science Association (NWO) via VIDI grant (452-17-004), and the Swiss National Science Foundation (SNSF) via a Ambizione grant (PZ00P3_174127). Data, code and replication instructions are available at OSF: `https://doi.org/10.17605/OSF.IO/TZNPY`. This study was approved by the Economics Business Ethics Committee at the University of Amsterdam (refs. 20170510120541 and 20210202020244) and preregistered on aspredicted.org (15709, 57718, 83830, 89876 and 124703).

Many common beliefs appear to be held for their comforting properties rather than their realism. Billions of adherents of the major religions believe in an afterlife, without concrete proof for its existence. Moreover, religiosity is higher in populations that face unpredictable shocks like earthquakes (Sinding Bentzen, 2019), during pandemics (Sinding Bentzen, 2021), and in the absence of alternative forms of insurance (Auriol et al., 2020). People at risk of serious diseases avoid medical testing and remain optimistic about their health status (Lerman et al., 1998; Oster et al., 2013; Ganguly and Tasoff, 2016), while greater exposure to Covid-19 leads people to become more sanguine about the probability of infection (Orhun et al., 2021; Islam, 2021). Populist politicians that promise easy fixes find more support in areas with weak economic prospects and declining growth rates (Mughan et al., 2003; Obschonka et al., 2018).

These findings are suggestive of wishful thinking, i.e. self-deception that is driven by a desire to feel better about the future. However, self-deception and its drivers are hard to pin down in field data.[1] Meanwhile, laboratory studies have yielded at best mixed evidence for wishful thinking, with several studies failing to support the phenomenon (see Section I). Strikingly, while the field studies usually focus on negative outcomes, the lab studies focus on positive ones, raising the question of whether wishful thinking is more prevalent in situations where people face potential losses and experience emotions such as fear and anxiety.

To better understand the link between adverse future outcomes, anticipatory anxiety and wishful thinking, we conduct a set of tightly controlled experimental studies. In our first four preregistered experiments (combined $N = 1,114$), we incentivize participants to correctly identify which of two types of patterns they see on their screen and induce anxiety by associating one type of pattern with an adverse outcome that may occur after a short waiting period. In our first experiment, the adverse outcome is a mild electric shock. In our other experiments, the adverse outcome is a monetary loss. Since participants have no control over the occurrence of these outcomes, the payoff-maximizing strategy is to simply identify the patterns as accurately as possible. By contrast, anticipatory anxiety about the shock or loss may cause wishful thinking, a belief that the anxiety-inducing state of the world is less likely than it really is. Consequently, wishful thinkers will be less accurate when the pattern that is flashed on the screen is associated with a shock or monetary loss, and more

[1]For instance, consistent with wishful thinking, Oster et al. (2013) find that people at risk of Huntington disease are optimistic before they get tested for the disease, but are reluctant to test, especially when they have low objective risk. However, without exogenous variation in the motives to hold optimistic beliefs, it is not clear whether initial optimism is the result of wishful thinking nor whether it is driven by a desire to avoid feeling anxious. Furthermore, Islam (2021) finds that individuals self-deceive about the risk of a Covid infection in deciding whether to go to a coffee shop during the pandemic. At the same time, they distort beliefs about the risk for others rather than for themselves, suggesting that self-deception is driven by social motives rather than anxiety about one's own health.

accurate when the pattern that is not flashed is associated with a shock or loss.

We propose a simple model to clarify the properties of wishful thinking in our experimental setting. Following Bénabou and Tirole (2002) and Brunnermeier and Parker (2005), we suppose that an agent self-deceives to optimally trade off the anticipatory utility benefits from alleviated anxiety and the material costs stemming from incorrect beliefs and subsequent decision-making. The model predicts that wishful thinking increases in the adversity of the outcome and the ambiguity of the evidence, and decreases with increased incentives for accuracy.

We find robust evidence for wishful thinking. In all of our first four experiments, participants are significantly less accurate in identifying patterns that may lead to an adverse outcome. This result obtains for different sources of anxiety (i.e. shock versus monetary loss), different pattern identification tasks, and in different environments (i.e. online versus laboratory). Ambiguous evidence facilitates wishful thinking across three different visual inference tasks with different manipulations of ambiguity. Wishful thinking remains high in later trials of the experiments, indicating its persistence. Because participants go through many trials, we can compute individual-level measures of wishful thinking to study heterogeneity in people's proclivity to engage in motivated cognition, a novelty in the experimental literature on this topic. We find that wishful thinking is stable within individuals, but heterogeneous across them. Finally, our dataset provides evidence against competing explanations for the observed phenomenon like an illusion of control, whereby participants believe that the pattern they report determines the adverse outcome, or the idea that adverse outcomes scare participants into providing noisy responses.

A key question in the literature on motivated beliefs is whether self-deception responds to the costs and benefits of holding biased beliefs. To investigate this, we manipulate the (material) costs of false beliefs by varying the accuracy bonus that participants can earn for a correct answer by factors of up to 200. In our first three experiments, higher accuracy incentives do not lead to a decrease in wishful thinking. They also do not lead to an increase in accuracy, despite an increase in response times and self-reported concentration. When we vary the psychological benefits of wishful thinking by manipulating the magnitude of monetary losses, we find that higher losses increase self-reported anxiety but have no statistically significant effect on wishful thinking. These results do not support the idea that self-deception takes into account material or anticipatory payoffs at the margin.

Next, we test for an alternative mechanism by which accuracy incentives may affect wishful thinking: higher incentives may induce additional effort to form accurate mental representations of patterns, thereby constraining wishful thinking much like the lower ambiguity of easier patterns does (see Online Appendix D.D for a model of this mechanism). Our first three experiments can

not provide a test of this mechanism because accuracy on the tasks is largely insensitive to cognitive effort. Instead, Experiment 4 features a self-timed pattern recognition task, where subjects can productively invest in gathering more information. In this setting, we find that higher incentives indeed reduce wishful thinking, specifically for those participants who increased their effort and accuracy. , Finally, several aspects of our data speak to the role of negative emotions like anxiety as a driver of wishful thinking. First, electric shocks are a well–established method to induce anxiety. Second, we verify that the size of monetary losses increases self-reported anxiety in our experiments. Third, self-reported anxiety is positively correlated with wishful thinking at the individual level. Lastly, in a fifth experiment ($N = 600$), we manipulate the framing of monetary outcomes as losses or gains. We replicate the finding of wishful thinking in the loss domain, but not in the gain domain, where we find evidence for pessimism, as in Huseynov et al. (2022). Thus, the anticipation of losses and its associated emotions appear to be a stronger driver of wishful thinking than the anticipation of gains. This may explain why previous laboratory experiments on wishful thinking, which have almost exclusively focused on the gain domain, have found mixed results.

In the next section, we review the experimental literature on wishful thinking and related phenomena. We then describe our experimental design. Section III introduces a simple theoretical model that helps us derive our hypotheses. Section IV contains the main results of our experiments, before we delve into the role of accuracy incentives (Section V) and anxiety (Section VI). Section VII provides a series of robustness checks of our main results. We conclude in Section VIII.

I Literature

People have been shown to self-deceive in the service of moral self-image (Kunda, 1990; Gino et al., 2016), ego utility (Eil and Rao, 2011; Möbius et al., 2022; Zimmermann, 2020), and a desire to be persuasive (Schwardmann and Van der Weele, 2019). At the same time, a small literature in experimental economics has investigated wishful thinking, i.e. self-deception motivated by anticipatory utility concerns, and failed to produce robust evidence. Unpublished work by Mayraz (2011) finds evidence for wishful thinking, but does not replicate in (Huseynov et al., 2022), who find the opposite tendency of apparent pessimism. Coutts (2019) finds evidence for wishful thinking in only one out of three tasks and Barron (2021) finds no evidence for asymmetries in updating of beliefs about the probability of winning monetary prizes. Mijović-Prelec and Prelec (2010) find evidence for wishful thinking in an experimental paradigm where wishful thinking could be confounded by confirmation bias.[2]

[2]Participants predict the type of a pattern before seeing it and are paid for their prediction. They are then also paid for identifying the pattern after seeing it, and their answer slants toward their prediction. The authors also find

The psychology literature on wishful thinking also features an active debate about the phenomenon's existence, scope and its underlying mechanisms. Some papers have studied wishful thinking by varying the desirability of one outcome over another. In a meta-analysis, Krizan and Windschitl (2007) find evidence for wishful predictions, but not for wishful thinking in confidence and subjective probability statements. Some papers on "motivated perception" are able to induce biased perceptions of ambiguous visual evidence (e.g., an image that could be interpreted as a B or a 13) by telling participants that one interpretation of the evidence results in the consumption of a preferred drink or food (Balcetis and Dunning, 2006). These studies struggle to rule out that participants believe that their answers can affect outcomes and they cannot incentivize beliefs because there is no true state of the world. Instead, they rely on implicit questionnaire items, eye tracking and reaction times to make extrapolations about participants' beliefs (Dunning and Balcetis, 2013). Leong et al. (2019) shows that monetary prizes affect visual perceptions and provides neurological evidence about the location of the perceptual distortions in the brain.

Our paper differs from the extant experimental literature by focusing on wishful thinking in the face of adverse outcomes.[3] In contrast to the paucity of evidence for wishful thinking derived from the gain domain, we find robust evidence for the phenomenon across perceptual tasks and sources of anticipatory utility. Moreover, we show experimentally that losses are special, by replicating the lack of wishful thinking in the gain domain. We are also able to isolate anxiety as a plausible driver of wishful thinking.

Our paper provides new insights on the role of accuracy incentives in disciplining wishful thinking, a central prediction of models of motivated beliefs (e.g. Bénabou and Tirole 2002; Brunnermeier and Parker 2005; Bénabou and Tirole 2011). Previous work by Armor and Sackett (2006) finds more optimism for hypothetical than for real events and Zimmermann (2020) shows that incentives can reduce motivated biases in recall. However, much evidence goes in the other direction. Simmons and Massey (2012) show that accuracy incentives of up to $50 do not correct football fans' overoptimistic expectations about their home team. Lench and Ditto (2008) find no effect of incentives on optimistic beliefs about adverse life events. Mayraz (2011) and Coutts (2019) find that higher rewards for accuracy do not reduce wishful thinking, and Schwardmann et al. (2022) find no evidence for an effect on self-persuasion and polarization in a debating context. Here, we

apparent wishful thinking in a control condition where the incentives for wishful thinking have been experimentally muted.

[3]Our focus on adverse outcomes connects to prior work that has claimed evidence for asymmetric updating about future life events, whereby bad news is downweighted (Sharot et al., 2011, 2012). But these results do not feature experimental variation and have been called into question, with critics suggesting that their results can be explained by standard Bayesian updating (Shah et al., 2016; Burton et al., 2022). Non-Bayesian asymmetric updating has been found in the domain of ego-relevant information, which may or may not capture anticipatory utility motives (Möbius et al., 2022). However, follow-up work has yielded mixed results (see Drobner, 2022, for a review).

find that accuracy incentives only reduce motivated beliefs in tasks where participants can improve the precision of signals through cognitive effort and thereby reduce the scope for wishful thinking. This suggests that the impact of economic incentives on motivated beliefs is likely to be highly sensitive to the nature of the inference task and the extent to which accuracy is elastic in effort.

We further contribute to the literature in two ways. First, we administer within-subject treatments with many observations per person, which allows us to show that wishful thinking is stable within individuals and differs between them.[4] Second, we vary the ambiguity of evidence in a subtle and inconspicuous way, allowing us to demonstrate that ambiguous evidence increases wishful thinking. This relationship between signal precision and wishful thinking replicates a robust finding in the previous literature on other forms of motivated beliefs (e.g. Haisley and Weber, 2010; Sloman et al., 2010; Chance and Norton, 2015; Gino et al., 2016; Grossman and Van der Weele, 2017) and helps explain how information avoidance can be an effective belief management tool.

II Design

Here we describe the design of Experiments 1 through 4, which we number in the order in which they were run. Experiment 5 is a simple variant of Experiment 2 and will be described in Section VI. We preregistered hypotheses for each experiment on `Aspredicted.org`. Preregistrations, IRB approvals and links to the experimental instructions can be found in Online Appendix E.

A Design features common to all experiments

In each experiment, participants engaged in a number of trials of a pattern recognition task. In each trial, they had to identify which of two possible types of pattern was shown on the screen. One of the two patterns was associated with the possibility of an undesirable outcome: an electric shock or a monetary loss, depending on the experiment. We refer to trials in which the pattern associated with a shock or loss and the pattern that was flashed on the screen were aligned as "shock/loss patterns" and trials in which they were not aligned as "no-shock/no-loss patterns".

If the no-shock/no-loss pattern was shown, then no shock or loss would occur in the trial. If a shock/loss pattern was shown on the screen, then the shock or loss occurred with a probability of one third at any point within an eight second period following the participants' response to the trial. This procedure injects objective uncertainty into the occurrence of the shock or loss. The probabilistic implementation also assures that shocks occur sparingly, which avoids rapid desen-

[4] Buser et al. (2018) do not find significant correlations between asymmetric updating of ego-relevant news across three tasks. However, in their study there are few observations per participant and the repeated updating task is noisy and subject to other biases like conservatism and base rate neglect.

sitization (or sensitization) of participants. Because participants will generally not be completely certain which pattern they saw, there is additional subjective uncertainty. In keeping with the previous literature, we will refer to the emotions induced by the threat of the shock or loss as "anticipatory anxiety".[5]

Our main treatment varies the associations between patterns and shocks or losses. Between trials and within participants, we varied not just the actual pattern but also which type of pattern was associated with a shock or loss. This assures that any differential response to the two types of patterns cannot affect our results. Moreover, the occurrence of the shock depended only on the pre-determined shock pattern and the actual pattern on the screen and not on a participant's response.

A participant who increases her subjective belief that she saw a no-shock pattern may reduce anxiety about the imminent shock or loss. She will also be less accurate in her response when a shock pattern is shown, and more accurate when a no-shock pattern is shown. This logic allows us to identify wishful thinking, which we measure as the difference between average accuracy for "no-shock" and "shock" patterns. Since accuracy is measured in percentage points from 0 to 100, wishful thinking can take values between 100 and -100. A value of 100 indicates maximum optimism, whereby a participant always guesses the no-shock pattern, whereas a value of -100 implies maximum pessimism.

Each experiment featured at least two further within-subject treatment variations. One of these varied the ambiguity of the pattern, in order to test whether wishful thinking is stronger for more difficult/ambiguous patterns. Another treatment varied the bonus that participants could win for a correct response, resulting in a *Low Accuracy Bonus* and a *High Accuracy Bonus* condition. This experimentally manipulated the trade-off between psychological payoffs from having more optimistic beliefs and the material payoffs from having more accurate beliefs. The order of these treatments was fully counterbalanced in each experiment. Participants received no explicit feedback about their performance.

Each experiment also implemented a series of variations on this basic structure in order to answer specific research questions. We summarize these variations in Table 1 and discuss each

[5]The American Psychological Association defines anxiety as "worry or apprehension about an upcoming event or situation because of the possibility of a negative outcome, such as danger, misfortune, or adverse judgment by others." The clinical psychology literature sometimes makes a distinction between fear and anxiety. Fear is defined as a behavioral response that serves to mobilize an organism in life-threatening situations that present immediate and identifiable danger. Anxiety, on the other hand, produces a more sustained response to aversive events that are unpredictable in terms of their timing and frequency, resulting in prolonged worry, tension and a feeling of insecurity (Grillon, 2008; Schmitz and Grillon, 2012). However, the fine points of the distinction differ between authors, and threats may induce a mixture of these emotions. Indeed, our design implements some elements of fear-induction (the threat is a clearly identifiable shock or loss) and anxiety-induction (the shock or loss is uncertain).

experiment in turn.

Table 1: Overview of experimental designs

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Participants	60	221	426	407
Number of trials	216	up to 96	up to 64	up to 96
Visual task	Single Gabor flash	Single Gabor flash	8 Gabor flashes	Colored dots
Anxiety source	Electric shock	Monetary loss	Monetary loss	Monetary loss
Loss/shock size	Self-calibrated	0, 0.1 or 5 pounds	1 pound	0 or 1 pound
Task difficulty levels	Tilt size (3 levels)	Tilt size (2 levels)	Likelihood ratio (continuous)	Dot ratio (4 levels)
Accuracy bonus levels	1 euro 20 euro	10 pence 8 pounds	5 pence 10 pounds	5 pence 10 pounds
Other design elements	Confidence measure Replication exp.		Treatment reminders	Treatment reminders Self-timed task
Start / end date	November 12, 2018 December 5, 2018	Feb 23, 2021 Feb 29, 2021	Jan 3, 2022 Jan 4, 2022	March 8, 2022 March 8, 2022
Location	CREED Laboratory (Amsterdam)	Online (Prolific.co)	Online (Prolific.co)	Online (Prolific.co)

B Experiment 1: Electric Shocks

The experiment took place in the CREED experimental laboratory at the University of Amsterdam. Sixty subjects were recruited from the CREED laboratory database, and participated in individual sessions. Upon coming to the lab, subjects read the instructions, signed a consent form and answered several control questions to determine their understanding of the task and the belief elicitation mechanism. The experimenter pointed out any wrong answers and discussed the correct answer until the participant indicated they understood them.

The source of anxiety in this experiment was a mild electric shock. Electric shocks are a proven method of inducing anticipatory anxiety.[6] Moreover, they are salient consumption events that afford a lot of control over the precise timing of the emotions. Since people differ in their pain thresholds, the strength of the electric shock was calibrated individually.[7]

[6]In particular, people pay to shorten the time they have to wait for electric shocks (Loewenstein, 1987; Berns et al., 2006) and they display physiological arousal while waiting for them, as reflected in a heightened skin conductance response (Grillon, 2008; Schmitz and Grillon, 2012; Engelmann et al., 2015, 2019).

[7]The wrist of the participant's non-dominant hand was connected to a Digitimer DS5 isolated bipolar current

The visual task was to determine whether a grating (Gabor patch), was tilted towards the left or right (see example in Panel (a) of Figure 1). Before each trial, subjects were reminded of the treatment conditions. After briefly seeing a fixation cross (750ms), the grating was flashed on the screen (150ms). Participants were then asked to indicate the direction of the tilt by pressing the left or right arrow on the keyboard (self-paced) as well as the confidence in their choice on a scale from 50% (completely uncertain) to 100% (certainty). We incentivized confidence ratings with a Becker-deGroot-Marschak (BDM) or "matching probabilities" mechanism. This mechanism makes it incentive compatible to state true beliefs, regardless of a particpant's risk preferences.[8]

Next, participants faced an anticipation screen (2000-8000ms), asking them to wait for the shock resolution. Finally, the electric shock was administered or not (1000ms). No trial-by-trial feedback was given about the correctness of the guess, but the average performance was communicated at the end of each block of 18 trials. Participants completed three sessions, each divided into four blocks of 18 trials. The four blocks correspond to four conditions of a 2x2 factorial design (Shock x Incentive). As described above, the Shock treatment varied whether the possibility of a shock was associated with a right-tilted or left-tilted grating pattern. The Incentive treatment varied whether the potential prize in the belief elicitation was 1 or 20 euros. We also varied the difficulty of the pattern recognition task within each block, by manipulating the degree of the tilt from the vertical line, where steeper patterns are harder to identify.[9]

Participants' earnings consisted of a 10 euro show-up fee, plus the earnings from the accuracy payments of one randomly drawn trial from both the low and high incentive condition. Thus, payments varied between 10 and 31 euros for a session that lasted on average slightly over an hour.

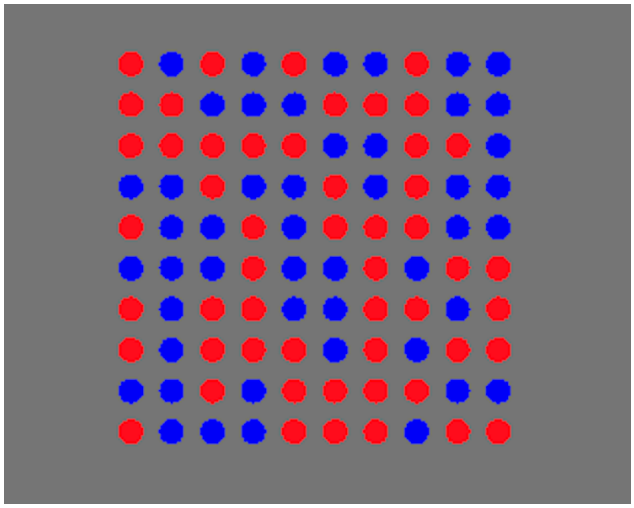
---

stimulator, which itself was connected to MATLAB through National Instruments USB x-series. The participant induced herself with a series of shocks, which she rated on a pain scale of 0 (not painful at all) to 10 (extremely painful). The calibration was complete when the subject rated the pain as 7-9 on the scale three consecutive times. A rating of 10 would lead to a decrease in the threshold. The maximum possible shock strength was set to 5V 25mA and the duration of the shock was set to 50ms (Engelmann et al., 2015, 2019).

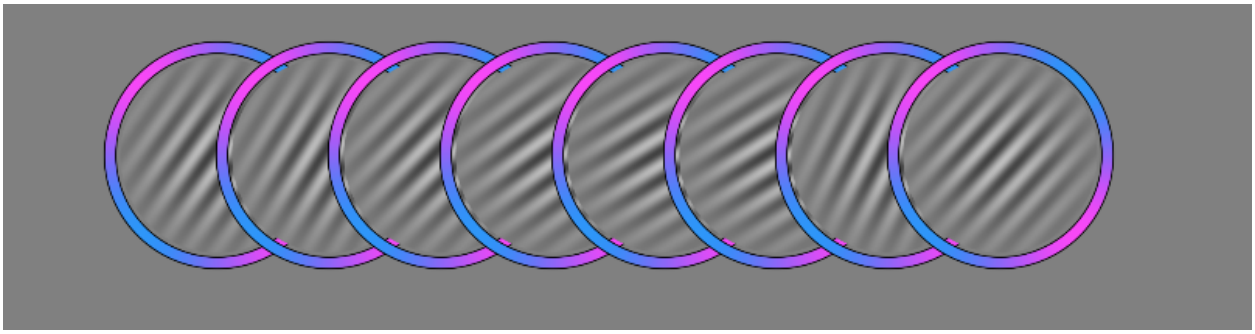
[8]Subjects indicate their subjective probability $x \in \{50, 55, ..., 95, 100\}$ that their answer was correct. The computer then randomly draws a number $z \in [50, 100]$. If $x \geq z$, then subjects win prize M, if their answer truly is correct. If $x < z$, then subjects win prize M with probability z. M varies between experimental conditions. Schlag et al. (2015) provide details about the origins and incentive compatibility of this mechanism, as well as evidence about its performance. After the instructions, but before the experiment started, participants had the opportunity to gain experience with the BDM mechanism.

[9]The three difficulty levels were calibrated to result in accuracy levels of 60%, 70% and 80%. Initially, these levels were calibrated on the basis of a pilot, and were the same for all subjects. To reduce the effects of fatigue or learning, difficulty levels were re-calibrated for each subject after each part, using a logistical performance function. This happened without subjects' knowledge, so this aspect of the design could not be gamed. We dropped the (re)calibration in the other experiments. We also had a few perfectly vertical trials that we drop in the analysis.

(a) Single gabor task (Experiment 1 and 2)



(b) Dot-counting task (Experiment 4)



(c) Multiple gabor task (Experiment 3)

Figure 1: Examples of the visual tasks in the various experiments.

C Experiment 2: Monetary losses as a source of anxiety

While electric shocks are a proven way to induce anxiety, they are not a common occurrence in everyday life. It is therefore important to understand whether the phenomenon carries over to other sources of anxiety, for instance the prospect of monetary losses.[10] Experiment 2 investigates wishful thinking in the presence of monetary losses. The experiment took place online, with 221 participants recruited from the online platform Prolific, which assures the highest quality of online data provision (Eyal et al., 2021). Participants had to answer a number of attention checks to advance to instructions, and a number of quiz questions about the instructions to advance to the experiment (see Online Appendix E). All monetary amounts were communicated in pounds.

To implement losses, participants were endowed with an amount of money and could lose part of this endowment in each trial. Participants were confronted with the same Gabor visual task as in Experiment 1. If a "loss pattern" appeared on the screen, then the participant would lose 20% of the endowment with a probability of one third. As before, subjects had to wait up to 8 seconds to learn whether they lost the money. To make losses salient, they were accompanied by an animation of an exploding bag of money. The experiment was divided into three parts of up to 32 trials. If the participant ran out of endowment before the 32 trials, then the remaining trials were cancelled.[11]

Using money allowed us to vary the size of the losses, and possibly the associated anxiety: Participants went through three parts of the experiment that varied in endowment and loss size: 25 pound endowment with 5 pound losses (*High Loss* condition), a 50 pence endowment with 10 pence losses (*Low Loss* condition), and no endowment with no threat of losses (*Neutral* condition). The Neutral condition served to address potential confounds that we discuss in Section VII.C, and was crossed with the treatments on accuracy incentives and difficulty.

To vary task difficulty, we used two different angles for the tilt of the pattern (see also Experiment 1). The accuracy incentives varied between trials to be either 8 pounds or 10 pence. Unlike in the previous experiment, we did not elicit confidence measures. Instead, we randomly selected one 8 pound trial and one 10 cent trial and paid subjects if their answer was correct. We made this change to implement the most parsimonious design that still allows for our various

[10] As we discuss in Section I, the connection between monetary outcomes and optimism has previously been investigated by other papers for positive sums of money, e.g. Mayraz (2011); Barron (2021); Coutts (2019), which has led to mixed findings.

[11] This design is informed by our conjecture that a slowly dwindling endowment is conducive to anxiety as losses accumulate irreversibly and the subjects see their (initially substantial) endowment slipping away. One alternative, to start each trial with a new endowment and pay one trial at random, may reduce anxiety as a) the loss will likely not count, b) the lost endowment is "replenished" immediately before the next trial. Whether this conjecture is true can be established by future research. Note that in rare cases, subjects ran out of money early enough that they had not yet experienced all possible treatments.

treatment dimensions, while avoiding attrition, fatigue and confusion of online participants due to the time-consuming and involved instructions of the confidence elicitations.

All treatments, including the three parts with different endowment sizes, were administered within-subject in randomized order. In order to reduce cognitive load, the tilt of the loss pattern (left vs. right) and the incentive for accuracy were varied at the block level, where a block consisted of 8 trials. At the start of each block, subjects were informed of the loss tilt, accuracy incentives and loss size, and were shown a reminder before the start of each individual trial. At the end of each block we conducted an interblock survey in which we asked participants for their agreement with two statements, measured on a five point Likert scale. The first stated that subjects were anxious to lose money from their endowment, the second that they were concentrated on the task.

D Experiment 3 and 4: Task characteristics and incentive effects.

Beside the source of anxiety, a second dimension of robustness concerns the visual decision-making task. The nature of the task matters for two reasons. First, if we are to take wishful thinking seriously as a cognitive phenomenon, it should be robust across multiple tasks, in contrast to evidence in Coutts (2019). Second, the task may affect mental trade-offs and hence the effect of accuracy incentives. In particular, incentives may reduce bias by motivating people to work harder to obtain evidence and thereby increase their accuracy, which then reduces their capacity for wishful thinking. Our quickly flashed Gabor pattern may not allow for increasing performance and may therefore not provide a good test of this mechanism.

To investigate these issues in more detail, we selected two new tasks that draw on more effortful cognitive processes. In doing so, we build on literature showing that the elasticity of performance to effort is task dependent (Camerer and Hogarth, 1999). To better test the effect of accuracy incentives, we reduced potential distractions in treatment variation by keeping loss sizes fixed. We also highlighted the accuracy incentive variation by alerting subjects explicitly that performance on high bonus trials was more lucrative.[12]

Experiment 3: Memory and inference task. The task in Experiment 3 is based on Drugowitsch et al. (2016) – see also Salvador et al. (2022). Participants saw a sequence of 8 tilted Gabor patches spaced over 4 seconds, as illustrated in Figure 1. The tilts were generated from one of two distributions of patterns that were biased towards either left or right-leaning patterns. We then asked participants to infer which distribution generated the patterns, and define a correct

[12]Instructions mentioned that "High Prize trials have a stronger impact on earnings than Low Prize trials. Participants who focus more on High Prize trials earn more on average than those who focus more on Low Prize trials."

answer as the one that corresponds to the distribution with the highest posterior likelihood given the displayed patterns.[13]

This task requires memorizing and mentally combining the several cues, which has been identified as a bottleneck of decision accuracy beyond the visual processing and choice implementation steps that were the focus of our previous task (Drugowitsch et al., 2016; Findling and Wyart, 2021; Wyart and Koechlin, 2016). It therefore requires a new dimension of mental effort, through which incentives for accuracy may increase decision accuracy and/or reduce bias. This design builds on evidence that incentive effects are larger for more complex tasks (Garbers and Konradt, 2014).

The design of the loss treatment followed that of Experiment 2. Participants completed two parts. In each part they received an endowment of 5 pounds from which they would lose 1 pound with a probability of one third if a "loss pattern" appeared. The part finished when the endowment was exhausted (after 5 losses) or after 32 trials. Within each part of the experiment, there were up to four 8-trial blocks across which we varied the size of the accuracy bonus (5 pence vs. 10 pounds) and the orientation of the loss patterns (left vs. right). After each block, there was an interblock survey that asked about concentration on the task (see Experiment 2). We recruited 426 subjects on Prolific, using the same procedures as in Experiment 2.

Experiment 4: Dot task. To further increase the link between mental effort and performance, we introduce a dot-counting task, displayed in Figure 1. Participants saw an array of 100 dots and were asked to identify whether the majority of dots were blue or red. The task was self-timed, with a time limit of 40 seconds. This allowed participants to exercise a lot of control over their performance through the time they spend on verifying the correct answer, including by counting the dots on the screen. Perhaps for that reason, previous studies using these or very similar tasks have found effects of incentives for accuracy (Caplin and Dean, 2014; Dean and Neligh, 2019; Dewan and Neligh, 2020). In addition, Bosch-Rosa et al. (2021) found evidence for motivated belief formation in this task.

The design followed that of Experiments 2 and 3. In each of two parts, participants received an endowment of 5 euros from which they would lose 1 euro with a probability of one third if a "loss pattern" appeared. A part finished when the endowment was exhausted (after 5 losses) or after 32 trials. Within each part of the experiment, there were up to four 8-trial blocks across which we varied the size of the accuracy bonus (5 pence vs. 10 pounds) and the color of the loss pattern (blue vs. red). We varied the difficulty of the task, by varying whether the majority color has 51, 52, 53 or 54 dots. In addition, we included one "Neutral" part of 32 trials without endowments

[13]Ocasionally, this might differ from the actual distribution that generated the pattern, but in contrast to Drugowitsch et al. (2016), we focus on the correct answer from the perspective of the participant.

or losses, the order of which was randomized to be either before or after the two parts with loss trials, and was crossed with the treatments on accuracy incentives and difficulty. Experiment 4 also featured the intertrial self-reports about anxiety and concentration that we used in Experiment 2. For Experiment 4, we recruited 407 participants on Prolific.

III Theoretical Predictions

In this section, we present a stylized model of wishful thinking that captures our experimental context and allows us to derive our main hypotheses. We will focus on the setting of experiment 1 and suppose that the threat of electric shocks is the source of anxiety. We assume that the agent chooses her beliefs trading of the anticipatory utility benefits of optimism with the material costs stemming from wrong decisions (Brunnermeier and Parker, 2005) and that belief distortions come at a cognitive cost (Bénabou and Tirole, 2002; Bracha and Brown, 2012).

The state of the world is given by $r_\theta \in \{0, 1\}$, where subscript θ refers to the true pattern and $r_\theta = 1$ means that it is *right-tilted*. A participant observes a pattern or visual signal s, and forms an initial probabilistic belief that $r_\theta = 1$, which we denote by $p(r_\theta, s) \in [0, 1]$. These undistorted initial beliefs $p(r_\theta, s)$ depend on the true state r_θ, with $p(r_\theta = 1, s) \geq 0.5$ and $p(r_\theta = 0, s) \leq 0.5$. They also depend on the precision of the visual signal, with $\frac{dp(r_\theta=1,s)}{ds} > 0$ and $\frac{dp(r_\theta=0,s)}{ds} < 0$. In particular, they become more certain when the signal is more precise.

After perceiving the pattern and forming her initial beliefs, the agent self-deceives into a new belief $\hat{p} \in [0, 1]$. Assuming that the agent states her chosen belief $\hat{p}$, the Becker-DeGroot-Marshak (BDM) mechanism implies the following expected material payoffs from potentially winning a prize M:

$$\pi(p, \hat{p}) = \frac{1}{2}\left(1 + 2p\hat{p} - \hat{p}^2\right) M$$

The probability of winning the prize is maximized at $\hat{p} = p$. Therefore, if material payoffs were the only object in the agent's utility function, then she would not self-deceive.[14]

The agent's anxiety of the electric shock is based only on her chosen beliefs $\hat{p}$ and is given by

$$\sigma_z(r_z\hat{p} + (1 - r_z)(1 - \hat{p}))qZ$$

The parameter $\sigma_z \geq 0$ captures the importance of anticipatory utility concerns, or a participant's

[14]The BDM mechanism was used in experiment 1 whereas experiments 2-4 paid participants for accurately identifying a given pattern. We cast the model in terms of the BDM mechanism for its analytical convenience. After Experiment 1 established similar results for confidence and (binary) accuracy judgments, we implemented discrete incentives in the subsequent online experiments in order to shorten instructions and reduce cognitive load.

innate anxiety. The parameter Z captures the utility loss due to a shock and q is the likelihood of a shock conditional on seeing a shock pattern. The parameter $r_z \in \{0, 1\}$ reflects whether the shock (hence, the subscript z) is associated with right-tilted ($r_z = 1$) or left-tilted ($r_z = 0$) patterns in a given trial. The agent will not only experience the disutility of anticipatory anxiety, but also the disutility of actually receiving the shock, which is given by $(r_z p + (1 - r_z)(1 - p))qZ$.

Suppose next that self-deception is not frictionless, but instead subject to a quadratic cognitive cost $\lambda(s)(p - \hat{p})^2$. The cognitive cost function is increasing in the distance between a participant's initial belief and her chosen belief. λ captures the magnitude of the cognitive cost and we assume that λ is increasing in s, the strength of the signal the agent encounters. Then, the agent's total utility is given by

$$\begin{aligned} U = \quad & \frac{1}{2}\left(1 + 2p\hat{p} - \hat{p}^2\right) M \\ & - (r_z p + (1 - r_z)(1 - p))qZ - \sigma_z(r_z \hat{p} + (1 - r_z)(1 - \hat{p}))qZ \\ & - \lambda(s)(p - \hat{p})^2 \end{aligned}$$

Maximizing the above expression with respect to $\hat{p}$ yields a participant's optimal belief

$$\hat{p}^* = p(s, r_\theta) - \frac{\sigma_z(2r_z - 1)qZ}{M + 2\lambda(s)}$$

From this optimal belief we can derive hypotheses about the effects of our experimental treatments. We consider the case in which the true pattern is right-tilted, $r_\theta = 1$, so that $\hat{p}$ is the belief in the correct answer. The case of $r_\theta = 0$ is symmetric. Then, the *Shock* condition corresponds to $r_z = 1$ and the *No-Shock* condition corresponds to $r_z = 0$. The amount of wishful thinking is given by

$$W := \hat{p}^*(r_z = 0) - \hat{p}^*(r_z = 1) = \frac{2\sigma_z qZ}{M + 2\lambda(s)} \tag{1}$$

From (1), and under the assumption that σ_z and λ are positive, we derive the following main hypothesis.

Hypothesis 1 (Wishful thinking) *There is positive wishful thinking, i.e.* $W > 0$.

Next, the effect of ambiguity on wishful thinking follows directly from our assumption that $\lambda'(s) > 0$.

Hypothesis 2 (Ambiguity) *Wishful thinking decreases when the pattern is easier to identify, i.e.* $\frac{dW}{ds} < 0$.

Our test of Hypothesis 2 illuminates how signal precision affects the production of distorted beliefs or a participant's ability to self-deceive. Signal precision s also affects $p(s, r_\theta)$, which in turn affects

the motivation to hold distorted beliefs. However, our symmetric design assures that $p(s, r_\theta)$ drops out of our measure of wishful thinking, allowing us to study participants' ability to self-deceive net of the strength of motives they may have to hold certain beliefs.

Next, the model predicts that higher accuracy incentives M raise the material costs of biased beliefs and make them less desirable.

Hypothesis 3 (Incentives) *Wishful thinking decreases in the size of the accuracy bonus, i.e.* $\frac{dW}{dM} < 0$.

Experiment 2 varies psychological stakes by varying the loss associated with a loss pattern. By relabelling Z to capture this monetary loss, we can state the following hypothesis.

Hypothesis 4 (Loss size) *Wishful thinking increases in the disutility of the adverse outcome, i.e.* $\frac{dW}{dZ} > 0$.

Online Appendix D features a number of extensions of the model. In Section D.A, we show that the predictions above are robust to also allowing the agent to derive anticipatory utility from her expectation of future accuracy payoffs. In Section D.B, we allow for a "bracing" or "defensive pessimism" motive for self-deception. We assume that, holding the actual likelihood of the shock constant, an agent suffers less disutility from the shock if she expects the shock to occur with a higher likelihood. Defensive pessimism works in the opposite direction of wishful thinking, so our main hypothesis can be rephrased as saying that wishful thinking trumps defensive pessimism as the dominant motive for belief distortion.

In Section D.C, we use the model to predict the correlation between measures of wishful thinking and (realized) anxiety, based on heterogeneities in fundamental parameters. We show that heterogeneity in λ implies a negative correlation and heterogeneity in σ_z implies a positive correlation.

Our data confirms some predictions of the model and is at odds with some others. To capture these discrepancies, section D.D proposes a revised model that allows for ex-ante investments in signal precision.

IV Main Results: Wishful Thinking

We start with an overview of the main results from our first four experiments. Table 2 shows OLS regressions of accuracy on our treatment variables. To deal with interdependence between observations for a given participant, we take as a unit of observation the average accuracy over an individual's trials within a given treatment and cluster standard errors at the participant level. Overall, 1,114 people participated in these four experiments, consisting of 48 percent females, and

with an average age of 34 (although the student sample in Experiment 1 is younger, with an average age of 21).

Our main hypothesis is that participants are less accurate in identifying patterns associated with a shock or monetary loss. Columns 1, 3, 5, and 7 of Table 2 exhibit strong evidence for such wishful thinking in each experiment. We see wishful thinking of 4.11 percentage points in experiment 1 ($p = 0.002$), 16.56 percentage points in experiment 2 ($p < 0.001$), 4.266 percentage points in experiment 3 ($p < 0.001$), and 8.453 percentage points in experiment 4 ($p < 0.001$).

We also hypothesize that wishful thinking is more pronounced for ambiguous or difficult patterns, where the signal is weaker and it may be easier to convince oneself of a positive outcome. The coefficient on the difficulty level across patterns shows participants are less likely to be correct on difficult patterns. The varying sizes of the coefficients across experiments reflect that difficulty levels were operationalized differently in the various experiments (see Table 1 for details). Crucially, the interaction terms in columns 2, 4, 6, and 8 show that the effect of loss or shock patterns increases with difficulty (all $p < 0.05$), thus confirming our hypothesis in all experiments.

Our third hypothesis is that incentives for accuracy reduce wishful thinking because they raise the costs of wrong beliefs. Table 2 shows no evidence for this hypothesis, as the interaction terms between loss/shock pattern and the accuracy bonus are not statistically significant (all $p > 0.1$). However, a closer examination in Section IV.C, reveals that accuracy incentives do have an effect in some settings. Finally, our fourth hypothesis is tested in column 4 of Table 2. We find that varying loss size, which we did in Experiment 2, has at most a small positive effect on wishful thinking that is not statistically significant.

If we average wishful thinking over all participants across the four experiments, then we find that average accuracy is 78.1 percent for no-shock/no-loss patterns and 69.8 percent for loss/shock patterns.[15] Therefore, the average effect of wishful thinking is 8.3 percentage points and seeing a shock/loss rather than a no-shock/no-loss pattern decreases performance above chance level by almost one third. These effect sizes are unlikely to be predictive of particular applications, as they show considerable context-dependence.[16]

[15] We use as an observation the individual averages of accuracy for shock/loss and no-shock/loss patterns, so that every individual is weighted the same regardless of the number of trials in the experiment she completed.

[16] It is hard to pinpoint the differences in effect sizes. Relative to Experiment 1, Experiment 2-4 replaced shocks with losses, but also took place online, which necessitated changes to the exact instructions, earnings and number of trials. A possible explanation for the smaller effect size in Experiment 1 is that because our recruitment message mentioned shocks, the experiment featured a selection of participants that was generally more comfortable with the relevant source of anticipatory anxiety. Experiments 3 and 4 further differ in the perceptual task and other implementation details.

Table 2: OLS regressions of accuracy levels on treatment across experiments

	Experiment 1 (Electric Shocks)		Experiment 2 (Monetary losses)		Experiment 3 (Repeat flash)		Experiment 4 (Dot task)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var:	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy
Shock/Loss Pattern	-4.111	-2.014	-16.54	-8.248	-4.266	-3.052	-8.452	-7.339
	(1.264)	(1.736)	(1.605)	(3.489)	(0.766)	(0.865)	(1.044)	(1.314)
High accuracy bonus (HAB)	0.785	0.313	-0.588	-1.081	0.630	0.685	1.732	1.050
	(0.878)	(1.387)	(0.851)	(1.089)	(0.474)	(0.601)	(0.628)	(0.856)
Difficult pattern (DP)	-8.602	-7.318	-15.68	-11.04	-20.55	-19.39	-7.064	-6.466
	(0.634)	(0.795)	(1.019)	(1.114)	(0.668)	(0.794)	(0.270)	(0.361)
Shock/Loss pattern x HAB		0.944		0.994		-0.110		1.363
		(1.787)		(1.771)		(0.881)		(1.325)
Shock/Loss pattern x DP		-2.569		-9.200		-2.317		-1.196
		(1.102)		(1.701)		(0.892)		(0.504)
Loss Size (LS)			-0.617	0.776				
			(0.906)	(1.245)				
Shock/Loss pattern x LS				-2.784				
				(1.869)				
Constant	80.75	79.70	85.82	81.65	87.66	87.06	89.53	88.98
	(1.106)	(1.287)	(1.964)	(2.310)	(0.791)	(0.829)	(0.734)	(0.800)
Observations	720	720	3415	3415	3408	3408	6502	6502
R^2	0.261	0.266	0.134	0.140	0.236	0.236	0.109	0.110

Notes: OLS regressions of accuracy on treatment dummies and interactions. Each observation is the average accuracy of an individual over all trials in a given treatment. "Shock/Loss pattern" is a dummy if the pattern is associated with a shock (Experiment 1) or loss (Experiments 2-4). "High accuracy bonus" is a dummy that represents a high accuracy bonus, while "Difficulty level" is a categorical variable that counts the difficulty level of the perceptual task, with the number of levels dependent on the experiment (see Table 1 for details). The continuous difficulty levels in Experiment 3 were binarized using a median split. "Loss Size" refers to the size of the monetary loss that we varied in Experiment 2. Standard errors in parentheses clustered by individual.

Nevertheless, as an external benchmark, one might consider mammogram reading, a complex pattern recognition task with a high-stake emotional outcome. Studies on interventions with radiologists often celebrate improvements in accuracy of a few percentage points, which are well in range of our effect sizes (Hadjiiski et al., 2004; Houssami et al., 2004).

Online Appendix A provides additional overviews and analysis, and shows that the regression results are robust to a panel data approach that uses the observations in all trials with and without individual fixed effects.[17] Next we elaborate on the results of the individual experiments and develop additional insights and interpretations.

A Experiment 1: Electric Shocks

Figure 2 shows the average accuracy levels from Experiment 1, split by shock and no-shock patterns. Each observation is the individual average over all trials in a given category, so $N = 60$ in each category. Panel (a) compares average accuracy between shock and no-shock patterns, demonstrating wishful thinking of about four percentage points (72.3 vs. 68.6 percent). In Online Appendix C, we describe a replication of this main treatment effect in Experiment 1 with $N = 50$.

Panel (b) of Figure 2 displays the impact of the three difficulty levels, as defined by the size of the tilt of the pattern. There appears to be some wishful thinking for easy patterns (2.4 percentage points) and medium patterns (2.5 percentage points). However, Table A.4 provides interaction terms for each of the difficulty levels, and shows that wishful thinking is statistically significant only for the most difficult patterns, where it rises to about 8 percentage points. Finally, panel (c) displays the impact of raising the prize for the BDM mechanism from 1 to 20 euro. Wishful thinking is about 1.4 percentage points more pronounced under the low bonus than under the high bonus, but the difference between the two conditions is not statistically significant.

Confidence measure. In addition to the accuracy measure, we elicited a measure of confidence in having correctly identified the pattern, incentivized with a BDM mechanism. This allows us to construct a continuous measure of participants' perceptions: the variable "Belief" measures the subjective belief in the correct answer on a scale from 0 (meaning the subject indicated 100% confidence in the wrong answer) to 100 (meaning the subject indicated 100% confidence in the correct answer). Figure A.1 and Table A.5 in Online Appendix A show results for this belief variable that are analogous to those for accuracy. We find the effects for accuracy and confidence

[17] In addition, Online Appendix Table A.1provides descriptive statistics of accuracy levels for all of our experiments. Online Appendix Figure A.4 provides an overview of the cumulative distribution functions of accuracy in shock/loss and no-shock/no-loss patterns.

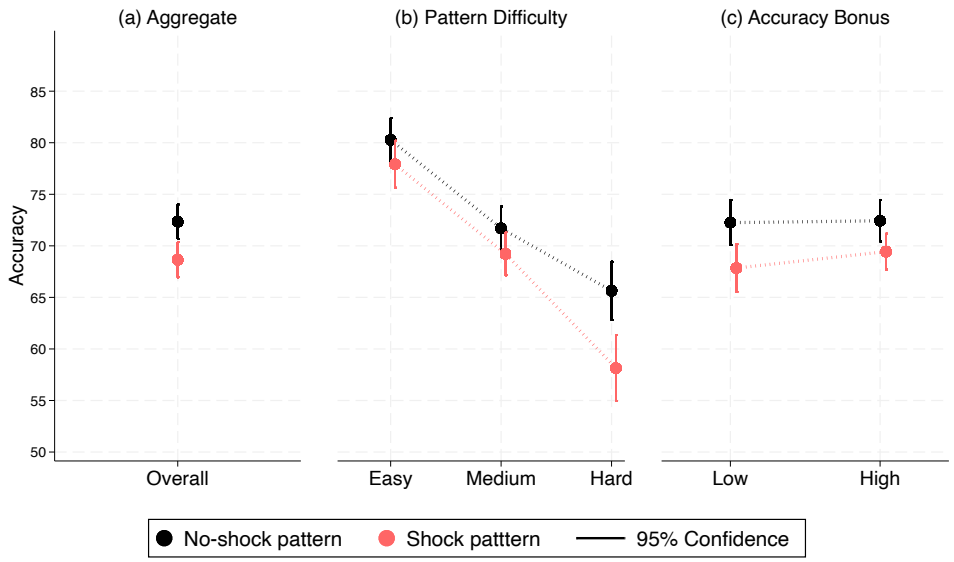


Figure 2: Electric shocks and accuracy in Experiment 1.

Notes: Average accuracy levels, split by shock and no-shock pattern. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category, so $N = 60$ in each category. Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty (tilt) of the pattern. Panel c) disaggregates by incentives for accuracy.

are comparable both in size and in statistical significance.[18]

B Experiment 2: Monetary losses as a source of anxiety

Experiment 2 replaced electric shocks with monetary losses. While the literature has documented how the threat of electric shocks increases anxiety, no such evidence is available for losses. As a manipulation check, we therefore asked subjects to report their agreement with the statement "I felt anxious about losing money from my endowment" on a scale from 1 to 5 after each treatment block of 8 trials in which losses could occur. Figure 3 shows the density of different anxiety ratings in the Low Loss (10 pence) and High Loss (5 pounds) condition. We find that average anxiety is 3.39 in the Low Loss condition and 4.15 in the High Loss condition ($p < 0.001$ on a linear regression with standard errors clustered by participant) and that participants report substantial levels of anxiety about monetary losses even in the Low Loss condition.

Turning to the main results, Figure 4 shows the average accuracy levels from Experiment 2, split by Loss and No-loss patterns. Each observation is an individual's average over all trials in a

[18]It is possible to conceive of the self-deception we see in our experiments as a Blackwell experiment inside the decision-maker's mind, where patterns associated with a shock or loss generate different rates of false positives and false negatives than patterns not associated with a shock or loss. A prediction of this interpretation is that the average belief of having seen a shock pattern is equal to the average prior. Our data suggest that this is not the case. Specifically, in Experiment 1, participants' average belief that they saw a shock pattern is 48.57 percent, which is biased away from the prior and true rate of 50 percent ($p<0.01$, t-test).

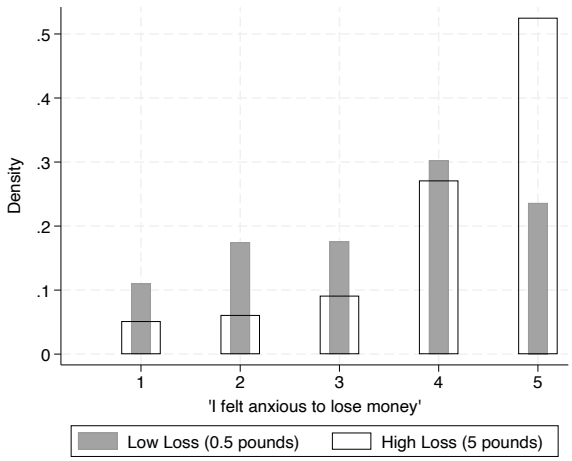


Figure 3: Manipulation check.

Notes: Histogram of agreement with the statement "I felt anxious about losing money from my endowment" measured on a 5-point Likert scale, split by loss size. Each report in a treatment block counts as one observation.

given category, so $N = 221$ in each category. Table 2, columns 3 and 4, provides regression evidence associated with these results, and Table A.6 provides robustness across regression models. Results exclude the Neutral condition, since this is not a test of wishful thinking and is discussed in Section VII.C.

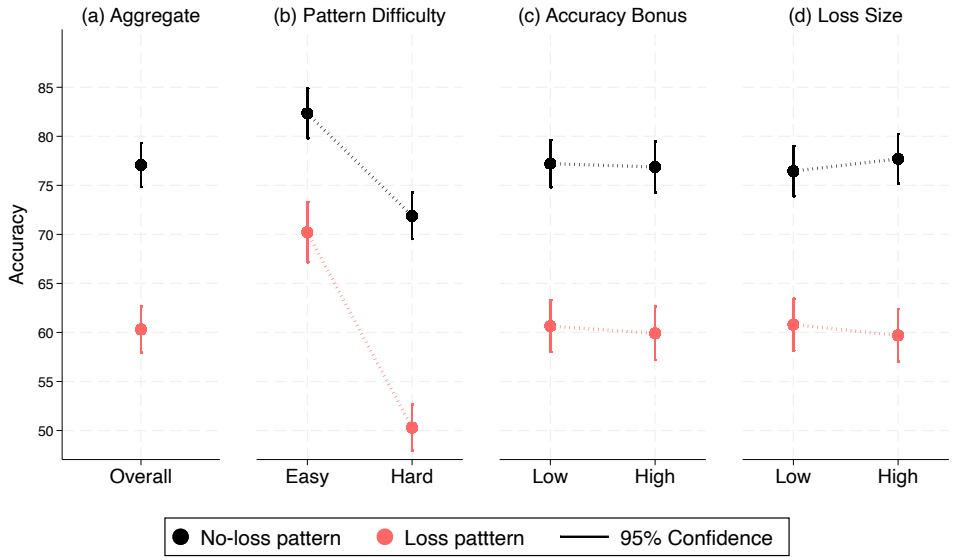


Figure 4: Monetary losses and accuracy in Experiment 2.

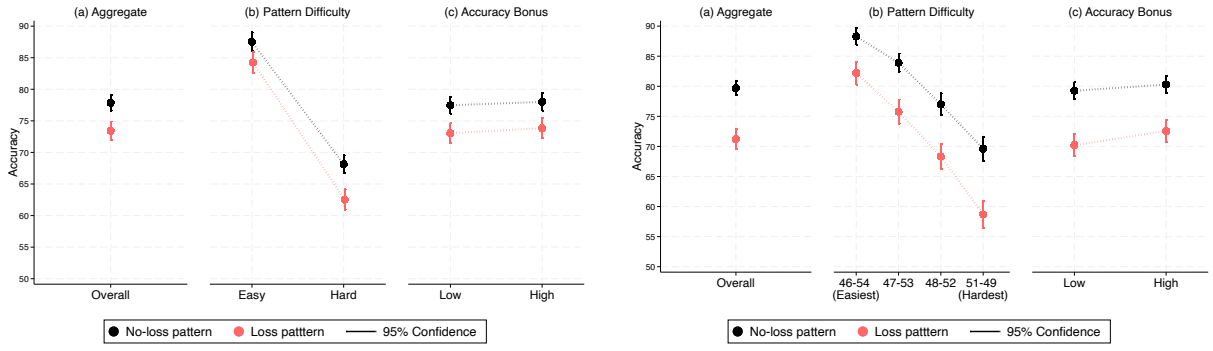
Notes: Average accuracy levels, split by loss and no-loss patterns. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category, so $N = 221$ in each category. Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty (tilt) of the pattern. Panel c) disaggregates by incentives for accuracy. Panel d) disaggregates by size of losses.

Panel (a) of Figure 4 compares average accuracy on No-loss and on Loss patterns. We see wishful thinking of 17 percentage points, which is highly statistically significant.The effect size is large: compared to the random-choice benchmark of 50 percent accuracy, accuracy is almost 3 times higher under patterns associated with no loss compared to those that are associated with a loss.

Panel (b) shows that there is wishful thinking for both pattern difficulty levels, as well as an interaction effect between wishful thinking and difficulty. Panel (c) shows the effect of seeing a loss pattern for accuracy bonuses of 0.1 and 8 pounds respectively. There is no evidence that incentives improve performance or reduce wishful thinking, which equals 16.6 percentage points under the low bonus and 17.0 percentage points under the high bonus. Panel (d) shows the effect of changing the loss size from 10 pence to 5 pounds. While this raises wishful thinking by about 2.7 percentage points, this difference is not statistically significant. Thus, the presence of losses can induce wishful thinking, but the size of losses does not affect the size of wishful thinking. This suggests the existence of some discontinuity in the effect of losses, which we discuss below and in Online Appendix D.

C Experiments 3 and 4: Task characteristics

Figure 5 shows the average accuracy levels in Experiments 3 and 4, split into the Loss and No-loss conditions. As before, each observation is the individual average over all trials in a given category.



(i) Sequential gabor task (Exp. 3)

(ii) Dot-counting task (Exp. 4)

Figure 5: Accuracy in the multiple gabor and dot-counting tasks in Experiment 3 and 4.

Notes: Average accuracy levels, split by loss and no-loss pattern. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category. In each subfigure, Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty of the pattern, with a median split shown for Experiment 3. Panel c) disaggregates by incentives for accuracy.

Figure 5.i shows the average accuracy levels from the sequential Gabor Task used in Experiment

3. Panel (a) compares average accuracy on no-loss patterns with the loss patterns, showing wishful thinking of 4.4 percentage points. Panel (b) displays the impact of task difficulty, which was a continuous variable in this task, defined by the posterior likelihood ratio of the two pattern-generating processes. The graph displays a median split on this variable, and shows a clear and statistically significant effect of higher difficulty on wishful thinking. Panel (c) shows wishful thinking for accuracy bonuses of 0.05 and 10 pounds respectively. Again, we find little evidence that incentives improve performance: A high bonus improves accuracy by about 0.7 ppt, but the effect is not close to being statistically significance. Moreover, there is no interaction with the loss pattern, so no reduction in wishful thinking from higher accuracy incentives.

Figure 5.ii shows the average accuracy levels from the Dot-counting task used in Experiment 4. Panel (a) shows wishful thinking of 8.5 percentage points in this task. Panel (b) displays the impact of pattern difficulty, where the easy patterns had a 46-54 split in colored dots, and the hardest patterns a 49-51 split. Once again, we confirm a statistically significant effect of difficulty on accuracy as well an interaction with wishful thinking. Panel (c) shows the pattern for the different levels of the accuracy bonus of 0.05 and 10 pounds. Unlike for the tasks we considered above, incentives improve performance: moving from the low to the high bonus improves accuracy by about 1.6 ppt ($p < 0.01$, see Table 2, column 8). On aggregate, we see at most a small interaction of accuracy incentives with the loss pattern. However, this result hides important heterogeneities between participants that we discuss in Section V.

D Dynamics

Our experiments consist of many trials and within-subject treatments, so we can ask how wishful thinking evolves over time. For instance, participants may get desensitized to the anxiety-inducing effects of electrical and monetary shocks and exhibit less wishful thinking in later trials. Or conversely, initial experiences with losses or shocks may heighten subsequent anxiety and increase wishful thinking in later trials. Such effects may offer a window into how motivated beliefs respond to experience and speak to mechanisms that may be at play in real world settings, which often feature dynamics and an element of repetition.[19]

Online Appendix A shows a number of analyses of the dynamics of wishful thinking over trials.[20]

[19]The presence of both anticipatory utility motives and a desire to avoid disappointment also has implications for the likely time-path of beliefs in the run-up to the realization of uncertainty (Macera, 2014). Unfortunately, our dataset only includes static beliefs.

[20]In Online Appendix Figure A.3 we provide a visual overview of wishful thinking over time in each experiment. Online Appendix Table A.14 analyses statistically how the effect of seeing a loss or shock pattern on accuracy (our measure of wishful thinking) evolves over time by interacting a dummy for whether a participant sees a loss pattern with the number of trials a participant has gone through. In a second set of analyses we simply compare wishful

Overall, the data do not present a coherent story, but suggest that repeated shocks may lead to desensitisation, while monetary losses can lead to heightened wishful thinking in some contexts. In particular, wishful thinking in the first half of the Experiment 1 is more than twice as large as in the second half. Although statistically non-significant ($p = 0.102$), this is suggestive of desensitization. By contrast, in Experiment 3, which features monetary losses, wishful thinking is higher in later trials and in the second half of the experiment. There is no significant effect of time or experience on wishful thinking in Experiments 2 and 4. Finally, Online Appendix Table A.15 investigates the effect of realized shocks or losses on wishful thinking in subsequent trials, but finds no effect in any of the four experiments, regardless of whether or not we control for a time trend in wishful thinking.

E Heterogeneity

Wishful thinking is usually identified by inducing experimental variation in participants' motives to hold biased beliefs. Since this experimental variation tends to be administered between subjects, the literature has not been able to obtain individual measures of a proclivity for motivated cognition and has therefore not been able to say much about individual differences (though see Buser et al., 2018). Our within-subject design with many trials allows us to explore individual differences.

Online Appendix Figure B.1 depicts histograms of individual-level wishful thinking in each experiment. It shows substantial variance, with a majority of participants engaging in some wishful thinking and some participants exhibiting the opposite effect. To establish that this apparent heterogeneity is not merely driven by measurement error or other sources of noise, we test for the stability of wishful thinking within individuals. In particular, we ask whether a participant's wishful thinking measured in one half of trials correlates with their wishful thinking in the other half. Online Appendix Table B.1 reports these half-split correlations of wishful thinking in Experiments 2, 3 and 4.[21] Correlations are around 0.5, with some fluctuations depending on how we split the data, indicating that heterogeneity in wishful thinking reflects individual differences. Wishful thinking is only slightly less stable than participants' skill in the pattern recognition tasks, as measured by the half-split correlations reported in columns 4 through 6 of Online Appendix Table B.1. To further show that these results are not driven by a few outliers, Online Appendix Figure B.2 shows the scatterplots pertaining to the odd-even trial splits in Table B.1.

Next, we correlate wishful thinking with a number of covariates of interest. First, we look at a

thinking in the first half and the second half of the experiment.

[21] We exclude Experiment 1 because there we recalibrated both the strength of the shock and the difficulty of the patterns during the experiment. This confounds the half-split correlations of wishful thinking and accuracy.

self-reported measure of concentration, which we measured in the interblock surveys of Experiments 2 and 4. Increased concentration may lead to more precise perceptions and higher accuracy, which in turn constrains wishful thinking, as we discuss in more detail in Section V. Second, we investigate self-reported "defensive pessimism", which measures the degree to which people adopt pessimistic beliefs to avoid disappointment.[22] This belief-based utility motive for self-deception into more pessimistic beliefs may arise if people are loss averse over changes in beliefs, as in Kőszegi and Rabin (2009). Defensive pessimism runs counter to wishful thinking, as we show formally in Online Appendix D.B, so one would expect a negative correlation.

Finally, we investigate the relationship between wishful thinking and self-reported anxiety about losing money from the endowment, which we measured in the interblock survey in Experiments 2 and 4. The sign of this correlation is theoretically ambiguous, as we explain formally in Online Appendix Section D.C. If the primary source of heterogeneity between participants is their proneness to anxiety, then wishful thinking should be positively correlated with experienced anxiety. Conversely, if participants vary strongly in their ability to self-deceive, then higher wishful thinking should be associated with lower experienced anxiety, as people who are very good at wishful thinking become more relaxed.

Table 3 shows OLS regressions of wishful thinking on these three explanatory variables. To generate maximal statistical power, we pool the data from all experiments in which the relevant explanatory variables were elicited. All regressions contain experiment dummies to control for differences in wishful thinking that are based solely on differences in the experimental context. Column 1 shows that wishful thinking is negatively correlated with the average self-reported concentration on pattern recognition. The correlation between wishful thinking and defensive pessimism in column 1 is negative and not statistically significant at conventional levels. In column 2 we add a participant's average self-reported anxiety to the regression model. The regression excludes Experiments 1 and 3, where we did not elicit an anxiety report. Anxiety is positively correlated with wishful thinking, but only statistically significant at the 10 percent level.[23] In columns 3 and 4 we do a robustness check on these correlations and exclude participants who reported difficulties following the

[22] Our measure is based on the defensive pessimism questionnaire (Norem, 2008). Following Lim (2009), we focus on the pessimism sub-scale, which measures agreement with the following statements: 1. I often start out expecting the worst, even though I will probably do OK. 2. I worry about how things will turn out. 3. I often worry that I won't be able to carry through my intentions. 4. I spend lots of time imagining what could go wrong. 5. I imagine how I would feel if things went badly. 6. In these situations, sometimes I worry more about looking like a fool than doing really well.

[23] We also elicited Beck Anxiety Inventory (BAI), a more general measure of anxiety that screens for, among other things, frequent physical symptoms of anxiety. BAI correlates with our measure of self-reported anxiety about incurring monetary losses in the experiment (corr=0.28, $p < 0.001$), thereby validating our more focused and tailor-made measure. However, perhaps unsurprisingly, the positive correlation between BAI and wishful thinking is not statistically significant.

Table 3: Emotional and cognitive covariates of wishful thinking.

Dep. variable:	(1) Wishful Thinking	(2) Wishful Thinking	(3) Wishful Thinking	(4) Wishful Thinking
Concentration	-2.899	-3.457	-3.979	-4.987
	(0.958)	(1.220)	(1.071)	(1.342)
Defensive pessimism	-0.608	-1.072	-0.911	-1.503
	(0.399)	(0.609)	(0.425)	(0.678)
Anxiety		1.550		1.950
		(0.825)		(0.890)
Constant	32.87	31.78	38.59	38.81
	(4.942)	(6.390)	(5.590)	(7.326)
Experiment dummies	✓	✓	✓	✓
Restrictions	None	None	Difficult instructions <4 of 7	Difficult instructions <4 of 7
Observations	1050	625	744	422
R^2	0.066	0.053	0.086	0.076

Notes: OLS regressions of wishful thinking on cognitive covariates. Data are from experiments 2, 3 and 4 in columns 1 and 3 and from experiments 2 and 4 in columns 2 and 4. Columns 3 and 4 only include participants with one of the three lowest scores on the question "How difficult did you find it to follow the instructions of this experiment?" measured on a 7-point Likert scale from very easy to very difficult. All regressions contain experiment dummies. Standard errors in parentheses.

instructions. Excluding such potentially noisy participants results in stronger correlations between wishful thinking and all covariates, including defensive pessimism.

These results allow us to sharpen our interpretations of wishful thinking. First, the negative correlation with concentration suggests that cognitive effort can constrain wishful thinking through its effect on accuracy. Second, the negative correlation with defensive pessimism suggests that belief-based utility motives that run counter to wishful thinking exist and can be detected in the cross-participant heterogeneity of belief biases. Since defensive pessimism is a self-reported survey scale, its correlation with wishful thinking suggests that people are at least somewhat conscious of their tendencies for probability distortion. Finally, the positive correlation with self-reported anxiety suggests that anxiety is a plausible driver of wishful thinking, that people differ in their innate anxiety, and that these differences are not (fully) overcome by their wishful thinking.

V The Effect of Accuracy Incentives on Wishful Thinking

Across our experiments, we find wishful thinking despite incentives for accuracy. To calculate the monetary cost associated with this stubborn wishful thinking, we can look at Experiment 2, which

featured the most wishful thinking of all experiments and hence provides an upper bound of these costs. We zoom in on trials with loss patterns, which mirror the many applications where the truth is scary.[24] Comparing accuracy on such loss patterns in the High Bonus condition with accuracy in a set of Neutral trials in which no losses were possible, implies an expected monetary cost from wishful thinking of about 87 pence.[25] This corresponds to roughly 10 minutes of work on the Prolific platform.

Wishful thinking thus persists despite meaningful costs. Moreover, as we have seen, it does not appear responsive to the size of these costs. In this section, we sharpen our interpretation of this null effect. At face value, it falsifies the idea, prominent in the literature, that self-deception takes into account a trade off between the psychological benefits and the material costs of biased beliefs at the margin. An alternative interpretation is that participants simply did not care about or notice variation in the accuracy bonus. We discuss these possibilities in turn.

A Do incentives for accuracy increase cognitive effort?

Our experiments contain several measures of cognitive effort that allow us to assess the effect of the accuracy bonus. The first measure is the accuracy of guesses. As documented earlier, accuracy responds to incentives only in Experiment 4. However, it is possible that the presence of losses distracted participants from the accuracy bonus. Therefore, we consider also the Neutral condition in Experiments 2 and 4 that does not feature the threat of a loss. Table 4 shows OLS regressions of the impact of the accuracy bonus on accuracy in Experiment 2 and 4, in both the conditions with losses present (columns 1 and 3) and the Neutral conditions (columns 2 and 4). Online Appendix Table A.2 reports raw means per treatment in the Neutral condition. In all cases, the bonus does not have a statistically significant effect on accuracy.

It is possible that participants are simply unable to improve their accuracy in some of our experiments, even if they care about the incentives. It is therefore instructive to look at more direct measures of effort. The first such measure is response time, which reflects how carefully subjects

[24] Ex-post, the symmetric nature of the task means that sometimes wishful thinking decreases accuracy (when losses are associated with the correct answer) and sometimes it increases accuracy (when losses are associated with the incorrect answer). As a result, averaged over all trials, the presence of losses does not decrease accuracy. This does not mean that wishful thinking is a money maximizing strategy from the subjective perspective of the agent. For an unbiased participant who is unsure which pattern she saw, self-deception always has negative expected value. This is true regardless of whether the bias pushes towards less accurate answers (for shock patterns) or more accurate answers (for no-shock patterns), because the agent's only way to distinguish between these is her (initial) subjective belief.

[25] In the High Accuracy Bonus condition, participants could earn 8 pounds if their answer in a randomly selected trial belonging to that category was correct. In that condition, accuracy for loss patterns was 60.3 percent. Accuracy in trials that rule out any wishful thinking was 71.2. So across trials, associating the true state of the world with an anxiety-inducing outcome lead to a 10.9 ppt decrease in accuracy and an expected loss of $0.109 * 8 = 0.87$ pounds. For the most ambiguous patterns, the decrease in accuracy is 12.8 ppt, and the expected loss is 1.02 pounds.

Table 4: Regressions of cognitive effort on accuracy bonus.

	Accuracy				Response time (log)		Concentration	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High accuracy bonus (HAB)	-0.0110	0.905	1.610	1.152	0.0433	0.159	0.145	0.204
	(0.821)	(0.905)	(0.574)	(0.690)	(0.0125)	(0.0179)	(0.0340)	(0.0229)
Experiment no.	2	2	4	4	2	4	2	4
Conditions	Losses-present	Neutral	Losses-present	Neutral	All	All	All	All
Observations	11396	7072	21114	13024	18468	34133	18468	34138
R^2	0.027	0.030	0.033	0.044	0.004	0.012	0.007	0.013

Notes: Regressions of measures of cognitive efforts on a dummy for the high accuracy bonus by experiment. Columns 1-4 show regressions on accuracy levels in experiment 2 (columns 1 and 2) and experiment 4 (columns 3 and 4). Columns 1 and 3 show results from the Losses-present conditions, columns 2 and 4 from the Neutral condition. Column 5 shows regressions across all conditions/experiments where the outcome variable is log response time in each trial, measured in milliseconds. Column 6 shows regressions across all Experiment 2-4 where the outcome variable is self-reported concentration. Concentration is measured as agreement with the statement "In the past 8 trials I was very concentrated on the task" on 5-point Likert scale. Regressions control for pattern difficulty and include a constant (not reported). Standard errors in parentheses are clustered by individual.

consider their answer.[26] Column 5 and 6 of Table 4 show that the accuracy bonus significantly increases logged response times, with a higher point estimate in Experiment 4. A final measure of effort is self-reported concentration: At the end of each 8 trial block in Experiments 2, 3 and 4, we asked participants to report their agreement with the statement "I was very concentrated on the task". Columns 7 and 8 of Table 4 show that a higher accuracy bonus leads to a significant increase in self-reported concentration. Online Appendix Table A.9 shows that the results for response times and concentration hold also in the remaining experiments.[27]

Taken together, these results indicate that participants care about and react to the accuracy bonus in every experiment, but raise performance only in Experiment 4, where the experimental task was chosen to be very elastic to cognitive effort. The fact that participants react to the accuracy bonus in every experiment allows us to sharpen our interpretation of the null effect of accuracy incentives on wishful thinking: participants' self-deceptive efforts do not take into account material incentives at the margin.

[26]Response times are routinely used in cognitive science and economics as a measure of cognitive effort (e.g. Bettman et al., 1990; Camerer and Hogarth, 1999; Enke et al., 2021) and are an important component of recent theories of decision making (Fudenberg et al., 2018; Alós-Ferrer et al., 2021; Clithero, 2018). Because of the highly skewed nature of the response time distribution, which may be sensitive to outliers, we look at the logarithm of response times as an outcome variable, which is measured in milliseconds (results for raw response times are similar).

[27]One concern about self-reported concentration is that within-subject variation is driven by experimenter demand. To investigate this, we focus on the first block of trials in each experiment, which differ in accuracy bonus between, but not within subject. While we lose a lot of power, we find that the results remain significant at the 5 percent level when we pool the experiments as well as for Experiment 2 by itself.

B Incentive effects and investments in signal precision

Having ruled out that self-deceptive efforts take material costs into account at the margin, we now turn to a second channel through which incentives may affect wishful thinking: whenever higher incentives spur cognitive effort that then leads to higher accuracy, this increase in signal precision may constrain participants' ability to self-deceive much like our exogenously-varied pattern difficulty did. While Experiment 4 sees a statistically significant increases in both cognitive effort and accuracy under higher accuracy incentives, we did not observe an overall reduction in wishful thinking. However, this result may hide some important heterogeneity. In particular, one may expect wishful thinking to go down only among participants who revealed an explicit effort to gather information by counting the dots.

We elicit this form of ex-ante information acquisition by simply asking participants in the post-experimental questionnaire whether they counted dots. We find that 9% of subjects replied "Always", 38% replied "Sometimes", and 53% replied "Never". These answers are not cheap talk, as they correlate with participants' response times. The participants in these three answer categories have mean response times of 14.4 seconds, 6.0 seconds, and 3.1 seconds respectively. Moreover, as Figure 6 shows, there are large differences in accuracy between counters and non-counters that cannot be the result of experimenter demand. Dot counters are also generally more responsive to accuracy incentives, both in terms of accuracy (see Online Appendix Table A.10) and (log) response time (Online Appendix Table A.11).

Given that we find clear effects of incentives among those who count the dots, the question becomes how this increased information gathering impacts wishful thinking. Figure 6 shows evidence for a reduction in wishful thinking among the Sometimes category. Online Appendix Table A.12 shows that this interaction is indeed significant at the 5 percent level for that category, and marginally significant for all counters. The Never category show a slightly negative and insignificant interaction. The variation in accuracy between the categories of counters provides further evidence that effort reduces the scope for wishful thinking: the Always counters are on average correct 88% of the time (vs. 73% for the Never counters). This leaves little scope for wishful thinking, which is indeed highly reduced and statistically not significant for this group. The idea that higher incentives affect wishful thinking by improving the quality of signals is also consistent with the fact, reported in Section IV.E, that higher self-reported concentration is predictive of lower levels of wishful thinking at the individual level.

How do these results relate to our model in Section III? The model predicts that a higher accuracy bonus reduces wishful thinking, by affecting the belief choice conditional on a given signal. Our evidence speaks against the literal mechanism assumed in the model– i.e. implicitly weighing

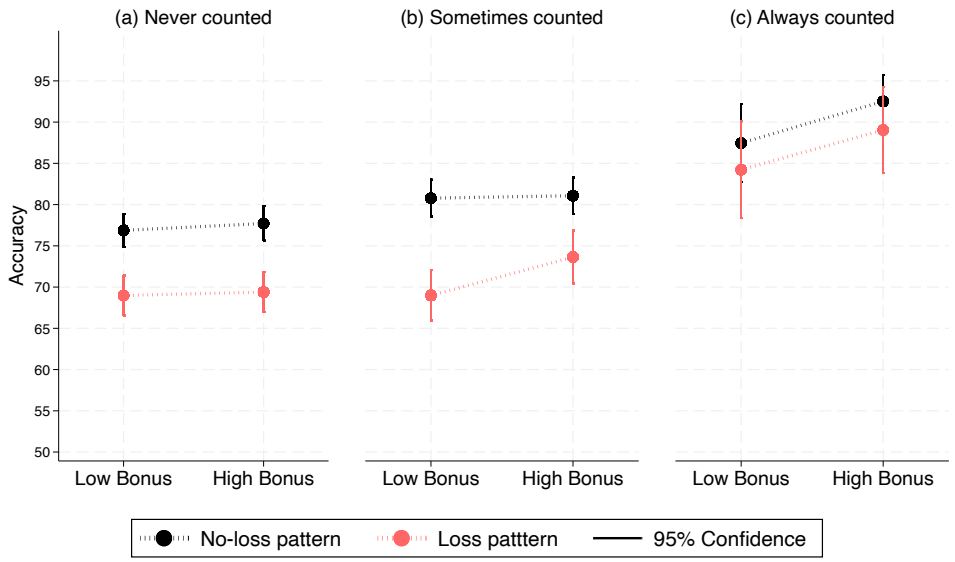


Figure 6: Accuracy in the dot-counting task.

Notes: Average accuracy levels, split by loss and no-loss patterns. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category. Panel a) shows participants who report that they never count ($N = 214$). Panel b) shows participants who sometimes counted ($N = 154$). Panel c) shows participants who always counted ($N = 36$).

the cost and benefits of self-deception. However, as we have shown, the general prediction that incentives constrain wishful thinking will still be correct in settings where people may gather more precise evidence.

In Online Appendix D, we propose an alternative model that accommodates all of our findings. We assume that self-deceptive efforts are costless up to a certain point, but impossible thereafter. This model implies that self-deception efforts are slow to respond to psychological and material incentives at the margin. However, successful investments in signal precision can constrain wishful thinking by improving signal quality and thereby lowering the maximum possible amount of self-deception. One interpretation of our results and the augmented model is that self-deception is closer to an "automatic" or "system 1" process that is constrained only by the precision of the signal (see also Kappes and Sharot, 2019; Melnikoff and Strohminger, 2020). In the augmented model, the effect of accuracy incentives does not necessitate that agents are sophisticated about the impact of signal precision on wishful thinking. What matters is that agents respond to incentives with a productive increase in cognitive effort. By contrast, the agent in the augmented model will only respond to higher exogenous losses or a greater motive to hold biased beliefs if they are sophisticated.

VI The Role of Losses and Anxiety: Experiment 5

Our experiments feature negative outcomes to generate anticipatory anxiety. We believe that anxiety is likely to be the dominant emotion in the case of electric shocks, because a previous literature has shown that shocks activate feelings of anxiety and fear (see Section II.B). The primacy of anxiety is also suggested by the correlation between individual measures of wishful thinking and self-reported anxiety that we document for Experiments 2 and 4 in Section IV.E. Experiment 2 further demonstrates that an increase in loss size increases self-reported anxiety.

In our fifth experiment ($N = 600$) we ask the related, reduced-form question of whether monetary losses are an especially strong driver of wishful thinking, as compared to gains. Our experiment features two conditions that differ according to whether outcomes are framed as losses or as (foregone) gains. As in Experiment 2, the patterns consisted of Gabor patches. In the Loss Frame treatment, subjects lost 50 pence if the loss pattern appeared on the screen from an initial endowment of 16 pounds. In the Gain Frame treatment, subjects gained 50 pence each time a gain pattern was flashed. Since there were 32 trials in each treatment, the distribution of outcomes is identical across treatments, with expected earnings of 8 pounds. This setup eliminates a layer of uncertainty compared to the other experiments, where losses occurred with a one-third probability conditional on the loss pattern. One trial was randomly selected for the payment of an accuracy bonus of one pound. After each block of 8 trials, we asked subjects about their experienced anxiety and their excitement about whether or not they would lose/gain money, measured on a 5 point Likert scale. We recruited 300 participants for each (between-subject) treatment on Prolific.co. The experiment was conducted in March 2023.

We hypothesize that the loss frame results in greater anxiety, less excitement and more wishful thinking (see the preregistration in Online Appendix E). We find the hypothesized treatment effects in self-reported emotions: in the loss domain we see higher self-reported anxiety, with an average individual score of 3.32, versus 2.88 in the gain domain ($p < 0.001$, t-test). For excitement, we see the reverse, with averages of 2.55 and 3.65 respectively ($p < 0.001$, t-test). Online Appendix Figure A.2 provides histograms of the distribution of reported emotions.

Figure 7 shows the results in terms of accuracy, with the corresponding average accuracy levels reported in Online Appendix Table A.3. Panel a) shows the aggregate results for all participants. Under the loss frame we observe wishful thinking of about 14 percentage points, replicating our previous results. Under the gain frame, we find a small and reversed effect, with participants being about 5 percentage points more accurate for no-gain patterns. Regression analysis in Online Appendix Table A.13 confirms the statistical significance of both these effects (column 2), as well as the overall significance of wishful thinking when combining the two domains (column 1).

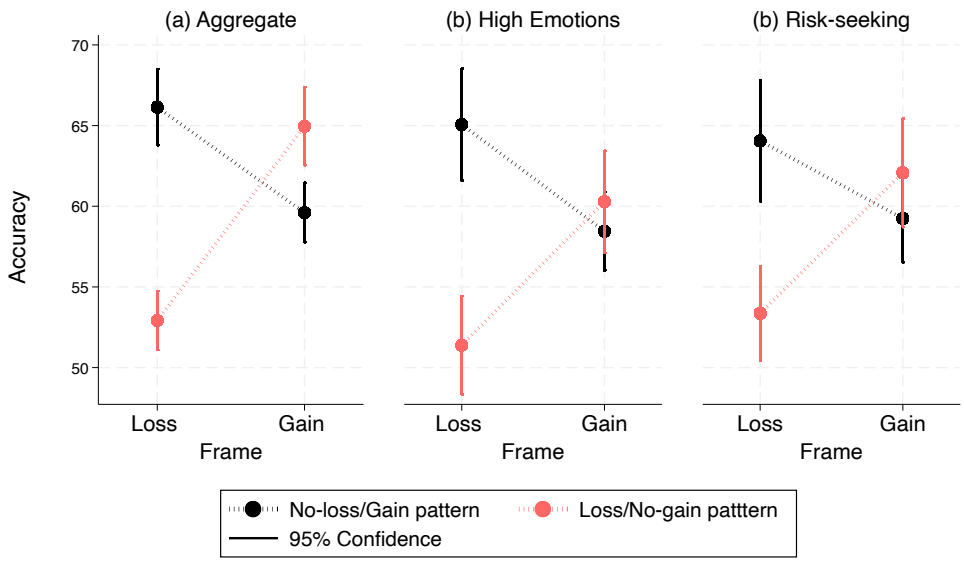


Figure 7: Accuracy in the Gain Frame and Loss Frame.

Notes: Average accuracy levels, split by loss (no-gain) and no-loss (gain) pattern. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category, so $N = 300$ in each category. Panel a) shows aggregate results. Panel b) shows only those subjects who self-report higher than 6 (an approximate median split) on an index that sums the average reported anxiety and excitement in the experiment. Panel c) shows only those subjects who self-report to be risk seeking on the question "Are you rather a risk-taking or risk-averse person (trying to avoid risks)?".

To further investigate the reverse effect in the domain of gains, we look at two moderators: (1) high levels of self-reported emotions, and (2) risk attitudes. We first look at the relationship with self-reported emotions. Panel b) of Figure 7 restricts the sample to participants who score above the median on an emotional index that sums average self-reports of anxiety and excitement, where the index is used to abate concerns over multiple hypothesis testing. In line with the idea that wishful thinking is more pronounced among those with higher anticipatory emotions, wishful thinking remains strong in the loss frame, but the reverse effect decreases and becomes statistically insignificant in the gain frame (see also Online Appendix Table A.13, column 3). This suggests that anticipatory emotions are still pushing perceptions in the direction of wishful thinking under the gain frame, but that they are less important than some other determinants of stated beliefs.

Panel c) looks into a potential explanation for the reverse effects. Participants may *hedge* by stating that they did not see a gain pattern. In the event of no-gain they are then partly insured by a potential payoff from the accuracy bonus. Such a strategy may be less salient in the loss domain, as it requires subjects to hedge across events of opposite valence (i.e. hedge losses against gains from accuracy). If hedging drives the reverse effect, then we would expect it to be less pronounced in risk-loving subjects. In line with this hypothesis, Panel c) of Figure 7 shows that the effect becomes small and statistically insignificant among those who self-report being strictly risk-loving

(see also Online Appendix Table A.13, column 4).

In sum, the results of experiment 5 further suggest that losses are special, plausibly because they inspire emotions of anxiety and fear. This helps explain why the previous experimental literature, which focused almost exclusively on the gain domain, has found little evidence for wishful thinking. Future research could further disentangle how the different emotions associated with gains and losses shape wishful thinking. Including physiological measures of anxiety in such research might be a particularly promising direction.

VII Robustness

At the end of each experiment, we asked participants several questions about their perceptions of the experiment and any potential confusion or mistrust they may have felt. In this section, we use these variables to conduct several robustness checks, where we pool the data from our four main experiments. We also test for an alternative interpretation of our results.

A Confusion and distrust

We check robustness by excluding various groups from our sample, one group at a time, namely participants who scored high on perceived difficulty of the instructions, who found it hard to recall the treatment conditions, who made more than two mistakes in the initial control questions, who did not trust the experimenters, or those whose accuracy in the experimental task was below 60 percent. The latter criterion excludes some participants who answer almost randomly and a small number of participants who almost always select the no-shock pattern.

The results are reported in Online Appendix Table A.16: wishful thinking remains highly significant in all selected samples, with small and statistically insignificant changes in effect sizes. The interaction of shock patterns with pattern difficulty also remains statistically significant in all specifications. The estimate for the interaction effect between the accuracy bonus and the shock pattern is generally positive but not statistically significant. Table A.17 shows similar results in analogous regressions where we use panel data from all trials and include individual fixed effects. We conclude that our results are not driven by misunderstanding or distrust.

B Illusion of control

Our experimental instructions stress that participants' answers do not have a causal effect on the shocks or losses. Several quiz questions during the instruction phase explicitly asked subjects to confirm their understanding of this point. Nevertheless, participants may have somehow come to

believe during the experiment that their answers were associated with shocks or losses. Such an "illusion of control" may lead subjects to switch their answers to the no-shock pattern.

To address this point, we conducted another understanding check in the closing questionnaire of Experiments 2, 3 and 4. A multiple choice question asked participants what drove losses in the experiment: a) the tilt of the pattern and designated loss category, b) their own answers, c) both, or d) don't know. On this question, the 81 percent of subjects who correctly gave the first answer had an average wishful thinking of 8.3 percentage points, while those who selected one of the other answers had average wishful thinking of 9.7 percentage points, a difference that is not statistically significantly ($p = 0.37$, t-test). In column 7 of Online Appendix Tables A.16 and A.17, we also run our main regressions without the participants who answered the control question incorrectly. We find that the estimated effect size for wishful thinking is statistically and quantitatively robust.

C Does seeing a shock or loss pattern increase noise?

It is possible that seeing a pattern that is associated with a possible loss or shock increases noise in participants answers, thereby reducing accuracy for shock patterns. This "noise-based explanation" supposes that participants perceive the correct answer initially, but that the anxiety from observing a shock pattern reduces performance through some form of interference that differs from wishful thinking.

This alternative account makes several predictions that we can test in the data. First, it implies a higher effect of the shock/loss threat for easier patterns, because these induce a higher subjective probability of seeing a shock pattern and should hence lead to higher noise. However, we see the reverse in the data. Second, the noise-based explanation predicts that average accuracy should increase in a neutral condition where there is no threat of a shock or loss at all. Performance in such an anxiety-free condition should exceed those on shock patterns as well as the aggregate performance under shock and no-shock patterns. Note that it need not be higher than the performance under no-shock patterns, as in this case self-deception goes in the direction of the correct answer and increases accuracy relative to neutral patterns.

To test this prediction, we use the Neutral condition in both Experiment 2 and Experiment 4. In one part of the experiment, implemented in random order, subjects were informed that they could not lose money from their endowment in any trial of this part. We compare accuracy for neutral patterns with accuracy for loss and no-loss patterns, where we pool the data from the two loss sizes in Experiment 2. As before, we take as an observation the individual accuracy rate in each of these conditions. In both Experiment 2 and 4, we find that average accuracy for neutral patterns is between that of the loss and no-loss patterns. Furthermore, there is not much evidence that stress

reduces average performance: in Experiment 2 accuracy is slightly (2.7 percentage points) higher in the Neutral condition than the average of the Loss and No-loss condition, but in Experiment 4 they are almost identical (see Online Appendix Table A.1 and A.2). Finally, a Neutral treatment in the replication of Experiment 1 further confirms these patterns, details of which are in Online Appendix C. We conclude that the data reject the noise-based explanation.

VIII Conclusion

Philosophers and economists have long considered the importance of beliefs for people's well-being. Jevons (1879) argues that "the greatest force of feeling and motive arises from the anticipation of a long-continued future", while Bentham (1789) points to expectation as being among the most significant sources of pleasure and pain. Over the last decades, economists have introduced anticipatory feelings as a source of utility into their formal models (Loewenstein, 1987; Caplin and Leahy, 2001) and the notion of utility from anticipation has experienced somewhat of a "renaissance" (Loewenstein and Molnar, 2018; Molnar and Loewenstein, 2021).

Our experiments show the importance of such anticipatory emotions for belief formation. In each of the four experiments, participants are significantly less accurate in identifying patterns that may result in adverse outcomes. Such wishful thinking is most pronounced when evidence is ambiguous, a result that replicates across tasks with distinct sources of ambiguity. Individuals differ in their propensity to engage in wishful thinking, with some showing the opposite tendency that reflects defensive pessimism. We find evidence that a higher material cost of wrong beliefs can reduce wishful thinking, but only when accuracy in the inference task is elastic to effort, so that participants can obtain more precise representations of signals if they choose to. Whether motivated beliefs respond to material incentives more generally is therefore likely to depend on the cognitive task and context in which beliefs are formed. Finally, we find that wishful thinking disappears in the domain of monetary gains, indicating that negative emotions are an important driver of the phenomenon.

Our findings speak to decision making in a wide range of applications, as anticipatory anxiety has been invoked in decisions related to health, insurance, finance and politics.[28] They help explain why people seek solace in religious beliefs, why financial professionals ignore red flags about their

[28]Examples are beliefs about health risks (Schwardmann, 2019), financial decisions (Brunnermeier and Parker, 2005; Bridet and Schwardmann, 2023), time inconsistency (Caplin and Leahy, 2001; Köszegi, 2010), occupational choice and the labor market equilibrium (Akerlof and Dickens, 1982; Santos-Pinto et al., 2018), information acquisition (Yariv, 2002; Eliaz and Spiegler, 2006; Loewenstein, 2006), principal-agent communication (Köszegi, 2006; Caplin and Leahy, 2004), self-image and taboos (Bénabou and Tirole, 2011), groupthink (Bénabou, 2013) and politics (Bénabou, 2008; Levy, 2014; Le Yaouanq, 2023).

asset portfolio, why people most at risk of a disease sometimes avoid testing for it, and why voters who are concerned about their jobs and the future of their children are susceptible to reassuring but false political narratives. The crucial role of ambiguity gives a rationale for the avoidance of precise information such as that provided in medical tests and helps explain the persistence of beliefs in phenomena such as the afterlife that, by their nature, do not admit clear evidence. Our findings on the role of accuracy incentives indicate that the bias can persist despite personal costs.

To further improve our understanding of wishful thinking it will be instructive to investigate other mediators of the phenomenon. Future experiments might explore whether wishful thinking is reduced by an opportunity to take a (costly) action to avert the adverse outcome that triggers it, and whether it responds to the length of the anticipation period. Moreover, our results on the role of accuracy incentives suggest that they operate mostly through increased information gathering. This raises interesting questions about the role of sophistication and the extent to which individuals design their informational environments to either facilitate or constrain their wishful thinking (Saccardo and Serra-Garcia, 2023).

References

Akerlof, George A. and William T. Dickens, "The economic consequences of cognitive dissonance," *The American Economic Review*, 1982, *72* (3), 307–319.

Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer, "Time will tell: Recovering preferences when choices are noisy," *Journal of Political Economy*, 2021, *129* (6), 1828–1877.

Armor, David A and Aaron M Sackett, "Accuracy, error, and bias in predictions for real versus hypothetical events.," *Journal of Personality and Social Psychology*, 2006, *91* (4), 583.

Auriol, Emmanuelle, Julie Lassebie, Amma Panin, Eva Raiber, and Paul Seabright, "God insures those who pay? Formal insurance and religious offerings in Ghana," *The Quarterly Journal of Economics*, 2020, *135* (4), 1799–1848.

Balcetis, Emily and David Dunning, "See what you want to see: Motivational influences on visual perception," *Journal of Personality and Social Psychology*, 2006, *91* (4), 612–625.

Barron, Kai, "Belief updating: does the 'good-news, bad-news' asymmetry extend to purely financial domains?," *Experimental Economics*, 2021, *24* (1), 31–58.

Bénabou, Roland, "Ideology," *Journal of the European Economic Association*, 2008, *6* (5), 321–352.

_ , "Groupthink: Collective Delusions in Organizations and Markets," *The Review of Economic Studies*, 2013, *80* (2), 429–462.

_ **and Jean Tirole**, "Self-confidence and personal motivation," *The Quarterly Journal of Economics*, 2002, *117* (3), 871–915.

_ **and** _ , "Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 2011, *126* (2), 805–855.

Bentham, Jeremy, *An Introduction to the Principles of Morals and Legislation*, Oxford: Clarendon Press, 1789.

Berns, Gregory S, Jonathan Chappelow, Milos Cekic, Caroline F Zink, Giuseppe Pagnoni, and Megan E Martin-skurski, "Neurobiological Substrates of Dread," *Science*, 2006, *754* (May), 754–758.

Bettman, James R, Eric J Johnson, and John W Payne, "A componential analysis of cognitive effort in choice," *Organizational behavior and human decision processes*, 1990, *45* (1), 111–139.

Bosch-Rosa, Ciril, Daniel Gietl, and Frank Heinemann, "Risk-Taking under Limited Liability: Quantifying the Role of Motivated Beliefs," *Working Paper, Available at SSRN 3985775*, 2021.

Bracha, Anat and Donald J. Brown, "Affective decision making: A theory of optimism bias," *Games and Economic Behavior*, 2012, *75* (1), 67–80.

Bridet, Luc and Peter Schwardmann, "Selling Dreams: Endogenous Optimism in Lending Markets," *Working Paper, Carnegie Mellon University*, 2023.

Brunnermeier, Markus K. and Jonathan A. Parker, "Optimal expectations," *American Economic Review*, 2005, *95* (4), 1092–1118.

Burton, Jason W, Adam JL Harris, Punit Shah, and Ulrike Hahn, "Optimism where there is none: Asymmetric belief updating observed with valence-neutral life events," *Cognition*, 2022, *218*, 104939.

Buser, Thomas, Leonie Gerhards, and Joël J. Van der Weele, "Responsiveness to Feedback as a Personal Trait," *Journal of Risk and Uncertainty*, 2018, *56*, 165–92.

Camerer, Colin F. and Robin M. Hogarth, "The effects of financial incentives in experiments: A review and capital-labor-production framework," *Journal of Risk and Uncertainty*, 1999, *19* (1), 7–42.

Caplin, Andrew and John Leahy, "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics*, 2001, *116* (1), 55–79.

_ **and** _ , "The supply of information by a concerned expert," *Economic Journal*, 2004, *114* (497), 487–505.

_ **and Mark Dean**, "Revealed preference, rational inattention, and costly information acquisition," *NBER Working Papers, (19876)*, 2014.

Chance, Zoe and Michael I Norton, "The what and why of self-deception," *Current Opinion in Psychology*, 2015, *6*, 104–107.

Clithero, John A, "Response times in economics: Looking through the lens of sequential sampling models," *Journal of Economic Psychology*, 2018, *69*, 61–86.

Coutts, Alexander, "Testing Models of Belief Bias: An Experiment," *Games and Economic Behavior*, 2019, *113*, 549–565.

Dean, Mark and Nate Leigh Neligh, "Experimental tests of rational inattention," *Discussion paper Columbia University*, 2019.

Dewan, Ambuj and Nathaniel Neligh, "Estimating information cost functions in models of rational inattention," *Journal of Economic Theory*, 2020, *187*, 105011.

Drobner, Christoph, "Motivated beliefs and anticipation of uncertainty resolution," *American Economic Review: Insights*, 2022, *4* (1), 89–105.

Drugowitsch, Jan, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin, "Computational precision of mental inference as critical source of human choice suboptimality," *Neuron*, 2016, *92* (6), 1398–1411.

Dunning, David and Emily Balcetis, "Wishful Seeing : How Preferences Shape Visual Perception," *Journal of Experimental Social Psychology*, 2013, *22* (1), 33–37.

Eil, David and Justin M Rao, "The good news-bad news effect: asymmetric processing of objective information about yourself," *American Economic Journal: Microeconomics*, 2011, *3* (2), 114–38.

Eliaz, Kfir and Ran Spiegler, "Can anticipatory feelings explain anomalous choices of information sources?," *Games and Economic Behavior*, 2006, *56* (1), 87–104.

Engelmann, J. B., F. Meyer, E. Fehr, and C. C. Ruff, "Anticipatory Anxiety Disrupts Neural Valuation during Risky Choice," *Journal of Neuroscience*, 2015, *35* (7), 3085–3099.

Engelmann, Jan B., Friederike Meyer, Christian C. Ruff, and Ernst Fehr, "The neural circuitry of affect-induced distortions of trust," *Science Advances*, 2019, *5* (3), 3413.

Enke, Benjamin, Uri Gneezy, Brian Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven, "Cognitive Biases: Mistakes or Missing Stakes?," *The Review of Economics and Statistics*, 2021, pp. 1–45.

Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, 2021, pp. 1–20.

Findling, Charles and Valentin Wyart, "Computation noise in human learning and decision-making: Origin, impact, function," *Current Opinion in Behavioral Sciences*, 2021, *38*, 124–132.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Speed, accuracy, and the optimal timing of choices," *American Economic Review*, 2018, *108* (12), 3651–84.

Ganguly, Ananda and Joshua Tasoff, "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future," *Management Science*, 2016, (September 2017), mnsc.2016.2550.

Garbers, Yvonne and Udo Konradt, "The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives," *Journal of Occupational and Organizational Psychology*, 2014, *87* (1), 102–137.

Gino, Francesca, Michael I Norton, and Roberto A Weber, "Motivated Bayesians: Feeling moral while acting egoistically," *Journal of Economic Perspectives*, 2016, *30* (3), 189–212.

Grillon, Christian, "Models and mechanisms of anxiety: evidence from startle studies," *Psychopharmacology*, 2008, *199* (3), 421–437.

Grossman, Zachary and Joël J Van der Weele, "Self-image and willful ignorance in social decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Hadjiiski, Lubomir, Heang-Ping Chan, Berkman Sahiner, Mark A Helvie, Marilyn A Roubidoux, Caroline Blane, Chintana Paramagul, Nicholas Petrick, Janet Bailey, Katherine Klein et al., "Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: an ROC study," *Radiology*, 2004, *233* (1), 255–265.

Haisley, Emily C and Roberto A Weber, "Self-serving interpretations of ambiguity in other-regarding behavior," *Games and economic behavior*, 2010, *68* (2), 614–625.

Houssami, Nehmat, Les Irwig, Judy M Simpson, Merran McKessar, Steven Blome, and Jennie Noakes, "The influence of clinical information on the accuracy of diagnostic mammography," *Breast Cancer Research and Treatment*, 2004, *85* (3), 223–228.

Huseynov, Samir, Mykel R Taylor, and Charles Martinez, "Farmers and Bakers: The role of Optimism Bias in Price Expectations," *Working Paper*, 2022.

Islam, Marco, *Motivated Risk Assessments*, Working Paper, Lund University, 2021.

Jevons, William S., *The Theory of Political Economy*, MacmIllan and Company, 1879.

Kappes, Andreas and Tali Sharot, "The automatic nature of motivated belief updating," *Behavioural Public Policy*, 2019, *3* (1), 87–103.

Köszegi, Botond, "Emotional Agency," *The Quarterly Journal of Economics*, 2006, *121* (1), 121–155.

_ , "Utility from anticipation and personal equilibrium," *Economic Theory*, 2010, pp. 415–444.

Kőszegi, Botond and Matthew Rabin, "Reference-dependent consumption plans," *American Economic Review*, 2009, *99* (3), 909–36.

Krizan, Zlatan and Paul D Windschitl, "The influence of outcome desirability on optimism.," *Psychological bulletin*, 2007, *133* (1), 95.

Kunda, Ziva, "The case for motivated reasoning," *Psychological Bulletin*, 1990, *108* (3), 480–498.

Le Yaouanq, Yves, "A model of voting with motivated beliefs," *Journal of Economic Behavior & Organization*, 2023, *213*, 394–408.

Lench, Heather C and Peter H Ditto, "Automatic optimism: Biased use of base rate information for positive and negative events," *Journal of Experimental Social Psychology*, 2008, *44* (3), 631–639.

Leong, Yuan Chang, Brent L Hughes, Yiyu Wang, and Jamil Zaki, "Neurocomputational mechanisms underlying motivated seeing," *Nature human behaviour*, 2019, *3* (9), 962–973.

Lerman, Caryn, Chanita Hughes, Stephen J. Lemon, David Main, Carrie Snyder, Carolyn Durham, Steven Narod, and Henry T. Lynch, "What you don't know can hurt you: Adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing," *Journal of Clinical Oncology*, 1998, *16* (5), 1650–1654.

Levy, Raphaël, "Soothing politics," *Journal of Public Economics*, 2014, *120*, 126–133.

Lim, Lena, "A two-factor model of defensive pessimism and its relations with achievement motives," *The Journal of Psychology*, 2009, *143* (3), 318–336.

Loewenstein, George, "Anticipation and the Valuation of Delayed Consumption," *The Economic Journal*, 1987, *97* (387), 666–684.

_ , "The Pleasures and Pains of Information," *Science*, 2006, *312* (May), 704–706.

_ **and Andras Molnar**, "The renaissance of belief-based utility in economics," *Nature Human Behaviour*, 2018, *2* (3), 166–167.

Macera, Rosario, "Dynamic beliefs," *Games and Economic Behavior*, 2014, *87*, 1–18.

Mayraz, Guy, "Wishful Thinking," *CEP Discussion Paper*, 2011.

Melnikoff, David E and Nina Strohminger, "The automatic influence of advocacy on lawyers and novices," *Nature Human Behaviour*, 2020, *4* (12), 1258–1264.

Mijović-Prelec, Danica and Drazen Prelec, "Self-deception as self-signalling : a model and experimental evidence," *Phil. Trans. of the Royal Society B*, jan 2010, *365* (1538), 227–240.

Möbius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat, "Managing self-confidence: Theory and experimental evidence," *Management Science*, 2022, *68* (11), 7793–7817.

Molnar, Andras and George Loewenstein, "Thoughts and players: An Introduction to old and new economic perspectives on beliefs," *The Science of Beliefs: A multidisciplinary Approach (provisional title, to be published in October 2021). Cambridge University Press. Edited by Julien Musolino, Joseph Sommer, and Pernille Hemmer*, 2021.

Mughan, A., C. Bean, and I. McAllister, "Economic globalization, job insecurity and the populist reaction," *Electoral Studies*, 2003, *22* (4), 617–633.

Norem, Julie, *The positive power of negative thinking*, Basic Books, 2008.

Norem, Julie K and Nancy Cantor, "Defensive pessimism: harnessing anxiety as motivation.," *Journal of personality and social psychology*, 1986, *51* (6), 1208.

Obschonka, Martin, Michael Stuetzer, Peter J. Rentfrow, Neil Lee, Jeff Potter, and Samuel D. Gosling, "Fear, Populism, and the Geopolitical Landscape: The "Sleeper Effect" of Neurotic Personality Traits on Regional Voting Behavior in the 2016 Brexit and Trump Elections," *Social Psychological and Personality Science*, 2018, *9* (3), 285–298.

Orhun, A Yesim, Alain Cohn, and Collin Raymond, "Motivated Optimism and Workplace Risk," *Working Paper*, 2021.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey, "Optimal expectations and limited medical testing: Evidence from huntington disease," *American Economic Review*, 2013, *103* (2), 804–830.

Pronk, Thomas, Dylan Molenaar, Reinout W Wiers, and Jaap Murre, "Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment," *Psychonomic Bulletin & Review*, 2021, pp. 1–11.

Saccardo, Silvia and Marta Serra-Garcia, "Enabling or limiting cognitive flexibility? evidence of demand for moral commitment," *American Economic Review*, 2023, *113* (2), 396–429.

Salvador, Alexandre, Luc H Arnal, Fabien Vinckier, Philippe Domenech, Raphaël Gaillard, and Valentin Wyart, "Premature commitment to uncertain decisions during human NMDA receptor hypofunction," *Nature Communications*, 2022, *13* (1), 1–15.

Santos-Pinto, Luís, Michele Dell, and Luca David Opromolla, "A General Equilibrium Theory of Occupational Choice under Optimistic Expectations," *Working Paper, University of Lausanne*, 2018.

Schlag, Karl H., James Tremewan, and Joël J. Van der Weele, "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs," *Experimental Economics*, 2015, *18* (3), 457–490.

Schmitz, Anja and Christian Grillon, "Assessing fear and anxiety in humans using the threat of predictable and unpredictable aversive events (the NPU-threat test)," *Nature Protocols*, 2012, *7* (3), 527–532.

Schwardmann, Peter, "Motivated health risk denial and preventative health care investments," *Journal of Health Economics*, 2019, *65*, 78–92.

_ **and Joel Van der Weele**, "Deception and self-deception," *Nature Human Behaviour*, 2019, *3* (10), 1055–1061.

_ **, Egon Tripodi, and Joël J van der Weele**, "Self-Persuasion: Evidence from Field Experiments at International Debating Competitions," *American Economic Review*, 2022, *112* (4), 1118–46.

Shah, Punit, Adam J. L. Harris, Geoffrey Bird, Caroline Catmur, and Ulrike Hahn, "A pessimistic view of optimistic belief updating," *Cognitive Psychology*, 2016, *90*, 71–127.

Sharot, Tali, Christoph W. Korn, and Raymond J. Dolan, "How unrealistic optimism is maintained in the face of reality," *Nature Neuroscience*, 2011, *14* (11), 1475–1479.

_ **, Marc Guitart-Masip, Christoph W Korn, Rumana Chowdhury, and Raymond J Dolan**, "How dopamine enhances an optimism bias in humans," *Current Biology*, 2012, *22* (16), 1477–1481.

Simmons, Joseph P and Cade Massey, "Is optimism real?," *Journal of Experimental Psychology: General*, 2012, *141* (4), 630.

Sinding Bentzen, Jeanet, "Acts of God? Religiosity and Natural Disasters Across Subnational World Districts," *Economic Journal*, 2019, *129* (622), 2295–2321.

_ , "In crisis, we pray: Religiosity and the COVID-19 pandemic," *Journal of Economic Behavior & Organization*, 2021, *192*, 541–583.

Sloman, Steven A., Philip M. Fernbach, and York Hagmayer, "Self-deception requires vagueness," *Cognition*, 2010, *115* (2), 268–281.

Wyart, Valentin and Etienne Koechlin, "Choice variability and suboptimality in uncertain environments," *Current Opinion in Behavioral Sciences*, 2016, *11*, 109–115.

Yariv, Leeat, "I'll See It When I Believe It - A Simple Model of Cognitive Consistency," *Working Paper, Available at SSRN 300696*, 2002.

Zimmermann, Florian, "The dynamics of motivated beliefs," *American Economic Review*, 2020, *110* (2), 337–61.

Online Appendix for "Anticipatory Anxiety and Wishful Thinking"

Jan B. Engelmann, Maël Lebreton, Nahuel A. Salem-Garcia,
Peter Schwardmann, Joël J. van der Weele

Table of contents

A Additional Tables and Figures

Table A.1: Average fraction of correct answers by experiment and treatment (excluding the Neutral condition).

	Experiment 1 (Elec. Shock) $N = 60$	Experiment 2 (Mon. Losses) $N = 221$	Experiment 3 (Seq. gabors) $N = 426$	Experiment 4 (Countable dots) $N = 407$
Aggregate	70.5	68.5	75.6	75.3
	(4.55)	(13.0)	(11.6)	(10.1)
No Shock/loss pattern	72.3	77.1	77.8	79.7
	(6.45)	(16.9)	(12.9)	(11.9)
Shock/loss pattern	68.6	60.3	73.4	71.1
	(6.57)	(17.5)	(15.3)	(16.4)
Difficulty Level 1	79.1	76.1	continuous	85.1
(easiest)	(6.17)	(17.6)		(12.8)
Difficulty Level 2	70.5	60.9	continuous	79.7
	(6.08)	(11.6)		(12.9)
Difficulty Level 3	62.9	-	continuous	72.6
	(7.95)	-		(14.6)
Difficulty Level 4	-	-	continuous	64.1
(hardest)	-	-		(15.0)
Accuracy bonus Low	70.1	68.7	75.2	74.7
	(5.68)	(14.2)	(12.6)	(11.4)
Accuracy bonus High	70.9	68.2	75.9	76.3
	(5.59)	(14.7)	(75.6)	(12.6)
Low Stake	-	68.5	-	-
	-	(14.0)	-	-
High Stake	-	68.7	-	-
	-	(15.0)	-	-

An observation is one individual's average accuracy in the specified condition. Standard deviations in brackets. Averages for Experiment 2 and 4 exclude the Neutral condition, which does not constitute a test of wishful thinking, and is reported in Table A.2.

Table A.2: Average fraction of correct answers by treatment in the Neutral condition.

	Experiment 2 $N = 221$	Experiment 4 $N = 407$
Aggregate	71.2 (19.1)	75.7 (18.2)
Difficulty Level 1 (easiest)	79.1 (20.1)	86.4 (13.8)
Difficulty Level 2	63.3 (14.1)	81.5 (15.2)
Difficulty Level 3	- -	72.3 (15.5)
Difficulty Level 4 (hardest)	- -	62.6 (18.1)
Accuracy bonus Low	70.8 (16.9)	75.1 (12.3)
Accuracy bonus High	71.7 (16.2)	76.3 (12.8)

An observation is one individual's average accuracy in the Neutral condition in Experiment 2 and 4 (data were not collected in Experiment 1 and 3). Standard deviations in brackets.

Table A.3: Average fraction of correct answers in the gain and loss treatments.

	Gain treatment $N = 300$	Loss treatment $N = 300$
Aggregate	62.2 (13.8)	59.7 (12.7)
No loss/Gain pattern	59.7 (16.2)	66.1 (20.7)
Loss/No gain pattern	64.9 (21.22)	52.8 (16.0)

An observation is one individual's average accuracy in the specified condition. Standard deviations in brackets.

A Experiment 1

Table A.4: Accuracy levels in Experiment 1

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy
Loss pattern	-3.698	-3.073	-3.698	-3.073	-4.111	-2.833
	(1.202)	(1.891)	(1.202)	(1.891)	(1.264)	(1.948)
High accuracy bonus (HAB)	0.885	0.174	0.885	0.174	0.785	0.313
	(0.857)	(1.270)	(0.857)	(1.270)	(0.878)	(1.389)
Medium difficulty (MD)	-8.606	-8.551	-8.611	-8.555	-8.639	-8.583
	(0.813)	(1.138)	(0.814)	(1.138)	(0.822)	(1.144)
High Difficulty (HD)	-17.15	-14.60	-17.14	-14.59	-17.20	-14.64
	(1.270)	(1.584)	(1.272)	(1.585)	(1.269)	(1.593)
Shock pattern x HAB		1.424		1.424		0.944
		(1.695)		(1.695)		(1.789)
Shock pattern x MD		-0.111		-0.111		-0.111
		(1.889)		(1.889)		(1.895)
Shock pattern x HD		-5.097		-5.097		-5.139
		(2.186)		(2.186)		(2.206)
Constant	80.50	80.19	80.50	80.19	80.76	80.12
	(1.020)	(1.250)	(0.871)	(1.085)	(1.056)	(1.314)
Observations	11520	11520	11520	11520	720	720
R^2	0.019	0.020	0.020	0.020	0.261	0.268
ID clustering	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓		
Averages by ID/treatment					✓	✓

Linear regressions of accuracy on treatment dummies. Columns 1-4 are panel data regressions, columns 3-4 include individual fixed effects. Columns 5-6 are OLS regressions where each observation is average accuracy per treatment and individual. Standard errors in parentheses clustered by individual.

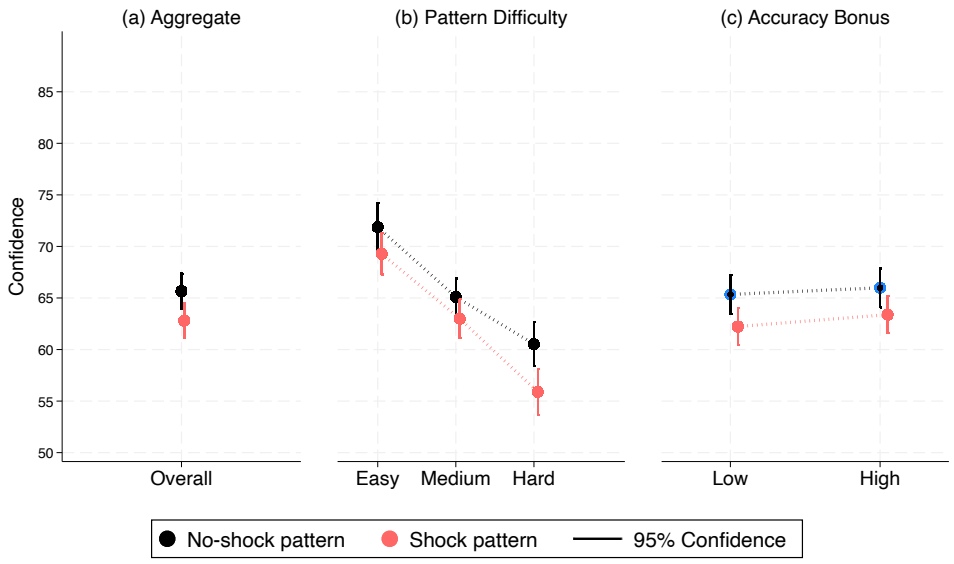


Figure A.1: **Electric shocks and confidence in the correct answer.** Average confidence levels in the correct answer, split by shock and no-shock pattern. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category, so $N = 60$ in each category. Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty (tilt) of the pattern. Panel c) disaggregates by incentives for accuracy.

Table A.5: Belief levels in Experiment 1

	(1) Belief	(2) Belief	(3) Belief	(4) Belief	(5) Belief	(6) Belief
Loss pattern	-2.858	-2.835	-2.858	-2.835	-3.108	-2.943
	(0.938)	(1.231)	(0.938)	(1.231)	(0.977)	(1.236)
High accuracy bonus (HAB)	0.898	0.642	0.898	0.642	0.841	0.478
	(0.533)	(0.788)	(0.533)	(0.788)	(0.562)	(0.842)
Medium difficulty (MD)	-6.503	-6.739	-6.516	-6.752	-6.532	-6.767
	(0.568)	(0.797)	(0.568)	(0.798)	(0.570)	(0.801)
High Difficulty (HD)	-12.34	-11.32	-12.32	-11.30	-12.37	-11.34
	(0.959)	(1.224)	(0.965)	(1.228)	(0.956)	(1.228)
Shock pattern x HAB		0.510		0.510		0.726
		(0.930)		(0.930)		(0.939)
Shock pattern x MD		0.472		0.472		0.470
		(1.109)		(1.109)		(1.112)
Shock pattern x HD		-2.040		-2.040		-2.056
		(1.348)		(1.348)		(1.360)
Constant	71.56	71.55	71.56	71.55	71.72	71.63
	(1.072)	(1.231)	(0.688)	(0.827)	(1.101)	(1.251)
Observations	11520	11520	11520	11520	720	720
R^2	0.027	0.028	0.028	0.029	0.255	0.258
Observations	11520	11520	11520	11520	720	720
R^2	0.027	0.028	0.028	0.029	0.255	0.258
ID clustering	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓		
Averages by ID/treatment					✓	✓

Linear regressions of beliefs on treatment dummies. Beliefs are constructed from confidence judgments on a scale from 0 to 100, where the latter is perfect confidence in the correct answer. Columns 1-4 are panel data regressions, columns 3-4 include individual fixed effects. Columns 5-6 are OLS regressions where each observation is average accuracy per treatment and individual and correspond to those in Table 2. Standard errors in parentheses clustered by individual.

B Experiment 2

Table A.6: Accuracy levels in Experiment 2

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy
Loss pattern	-16.59	-9.155	-16.59	-9.143	-16.54	-8.248
	(1.531)	(3.284)	(1.531)	(3.287)	(1.605)	(3.489)
High accuracy bonus (HAB)	-0.0284	-0.119	-0.0574	-0.157	-0.588	-1.081
	(0.806)	(1.004)	(0.812)	(1.012)	(0.851)	(1.089)
Difficult pattern (DP)	-15.29	-10.82	-15.29	-10.82	-15.68	-11.04
	(0.968)	(1.019)	(0.970)	(1.020)	(1.019)	(1.114)
Loss size (LS)	0.0380	1.100	0.0929	1.163	-0.617	0.776
	(0.893)	(1.212)	(0.892)	(1.209)	(0.906)	(1.245)
Loss pattern x HAB		0.221		0.242		0.994
		(1.618)		(1.620)		(1.771)
Loss pattern x DP		-8.806		-8.795		-9.200
		(1.586)		(1.586)		(1.701)
Loss pattern x LS		-2.079		-2.096		-2.784
		(1.813)		(1.815)		(1.869)
Constant	84.57	80.77	84.54	80.73	85.82	81.65
	(1.944)	(2.210)	(1.719)	(2.068)	(1.964)	(2.310)
Observations	11396	11396	11396	11396	3415	3415
R^2			0.064	0.066	0.134	0.140
ID clustering	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓		
Averages by ID/treatment					✓	✓

Linear regressions of accuracy on treatment dummies. Columns 1-4 are panel data regressions, columns 3-4 include individual fixed effects. Columns 5-6 are OLS regressions where each observation is average accuracy per treatment and individual and correspond to those in Table 2. Standard errors in parentheses clustered by individual.

C Experiment 3

Table A.7: Accuracy levels in Experiment 3

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy
Loss pattern	-4.415	-3.592	-4.407	-3.584	-4.266	-3.052
	(0.765)	(0.862)	(0.765)	(0.862)	(0.766)	(0.865)
High accuracy bonus (HAB)	0.728	0.484	0.723	0.479	0.630	0.685
	(0.471)	(0.589)	(0.473)	(0.590)	(0.474)	(0.601)
Difficult pattern (DP)	-20.76	-19.68	-20.80	-19.72	-20.55	-19.39
	(0.668)	(0.795)	(0.668)	(0.797)	(0.668)	(0.794)
Loss pattern x HAB		0.478		0.479		-0.110
		(0.838)		(0.838)		(0.881)
Loss pattern x DP		-2.137		-2.138		-2.317
		(0.849)		(0.849)		(0.892)
Constant	87.79	87.37	87.98	87.56	87.66	87.06
	(0.773)	(0.812)	(0.546)	(0.574)	(0.791)	(0.829)
Observations	33507	33507	33507	33507	3408	3408
R^2			0.064	0.064	0.236	0.236
ID clustering	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓		
Averages by ID/treatment					✓	✓

Linear regressions of accuracy on treatment dummies. Columns 1-4 are panel data regressions, columns 3-4 include individual fixed effects. Columns 5-6 are OLS regressions where each observation is average accuracy per treatment and individual and correspond to those in Table 2. Standard errors in parentheses clustered by individual.

D Experiment 4

Table A.8: Accuracy levels in Experiment 4

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy
Loss pattern	-8.438	-6.523	-8.455	-6.529	-8.452	-7.339
	(1.035)	(1.292)	(1.035)	(1.292)	(1.044)	(1.314)
High accuracy bonus (HAB)	1.729	1.205	1.798	1.286	1.732	1.050
	(0.567)	(0.801)	(0.572)	(0.807)	(0.628)	(0.856)
Difficult pattern (DP)	-7.023	-6.204	-7.025	-6.207	-7.064	-6.466
	(0.256)	(0.342)	(0.256)	(0.342)	(0.270)	(0.361)
Loss pattern x HAB		1.034		1.011		1.363
		(1.203)		(1.204)		(1.325)
Loss pattern x DP		-1.613		-1.613		-1.196
		(0.470)		(0.470)		(0.504)
Constant	89.37	88.40	89.31	88.33	89.53	88.98
	(0.707)	(0.794)	(0.674)	(0.757)	(0.734)	(0.800)
Observations	21114	21114	21114	21114	6502	6502
R^2			0.046	0.046	0.109	0.110
ID clustering	✓	✓	✓	✓	✓	✓
Individual fixed effects			✓	✓		
Averages by ID/treatment					✓	✓

Linear regressions of accuracy on treatment dummies. Columns 1-4 are panel data regressions, columns 3-4 include individual fixed effects. Columns 5-6 are OLS regressions where each observation is average accuracy per treatment and individual and correspond to those in Table 2. Standard errors in parentheses clustered by individual.

Table A.9: Regressions of cognitive effort on accuracy bonus across experiments

	(1) Concen-tration	(2) Concen-tration	(3) Concen-tration	(4) Response time (log)	(5) Response time (log)	(6) Response time (log)	(7) Response time (log)
High accuracy bonus (HAB)	0.117	0.120	0.173	0.0377	0.0525	0.0311	0.132
	(0.0330)	(0.0169)	(0.0259)	(0.0169)	(0.0129)	(0.00646)	(0.0177)
Experiment no.	2	3	4	1	2	3	4
Fixed effects				✓	✓	✓	✓
Observations	442	852	814	11520	11396	33507	21111
R^2	0.007	0.017	0.012	0.001	0.003	0.001	0.013

Regressions of cognitive efforts on a dummy for the high accuracy bonus by experiment. Columns 1-3 show regressions on self-reported concentration, where an observation is an individual's average concentration over all trials in the High Bonus and Low Bonus condition. Concentration is measured as agreement with the statement "In the past 8 trials I was very concentrated on the task" on 5-point Likert scale. Columns 4-7 show panel regressions where the outcome variable is log response time in each trial, where the latter is measured in milliseconds. Panel regressions include individual fixed effects. Standard errors in parentheses are clustered by individual.

Table A.10: Accuracy levels in Experiment 4 split by dot counting behavior

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
High accuracy bonus (HAB)	0.806	1.925	4.836	2.369
	(0.759)	(0.969)	(1.976)	(0.871)
Counting dots:	Never	Sometimes	Always	Sometimes/Always
Observations	11255	7839	1863	9859
R^2	0.000	0.001	0.006	0.001

OLS regressions of accuracy on the accuracy bonus, split by dot-counting behavior. Standard errors in parenthesis are clustered at the individual level.

Table A.11: Response times in Experiment 4 split by dot counting behavior

	(1) Response time	(2) Response time	(3) Response time	(4) Response time
High accuracy bonus (HAB)	0.0579	0.251	0.207	0.234
	(0.0165)	(0.0390)	(0.0750)	(0.0359)
Counting dots:	Never	Sometimes	Always	Sometimes&Always
Observations	11255	7836	1863	9856
R^2	0.002	0.019	0.013	0.014

OLS regressions of log response times on the high accuracy bonus, split by dot-counting behavior in different columns. Standard errors in parenthesis are clustered at the individual level.

Table A.12: Accuracy levels in Experiment 4 split by dot counting behavior

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
Loss pattern	-7.937	-11.46	-3.732	-10.12
	(1.733)	(2.090)	(2.849)	(1.810)
High accuracy bonus (HAB)	1.220	-0.171	4.389	0.720
	(1.097)	(1.356)	(2.306)	(1.177)
Loss pattern x HAB	-0.795	4.131	0.898	3.263
	(1.721)	(1.950)	(3.021)	(1.660)
Constant	76.69	80.64	87.85	81.88
	(0.991)	(1.157)	(2.209)	(1.048)
Counting dots:	Never	Sometimes	Always	Sometimes/Always
Observations	11255	7839	1863	9859
R^2	0.009	0.013	0.008	0.012

OLS regressions of accuracy on treatment dummies, split by dot-counting behavior. Each observation is the average accuracy per treatment and individual. Standard errors in parenthesis are clustered at the individual level.

E Losses versus gains

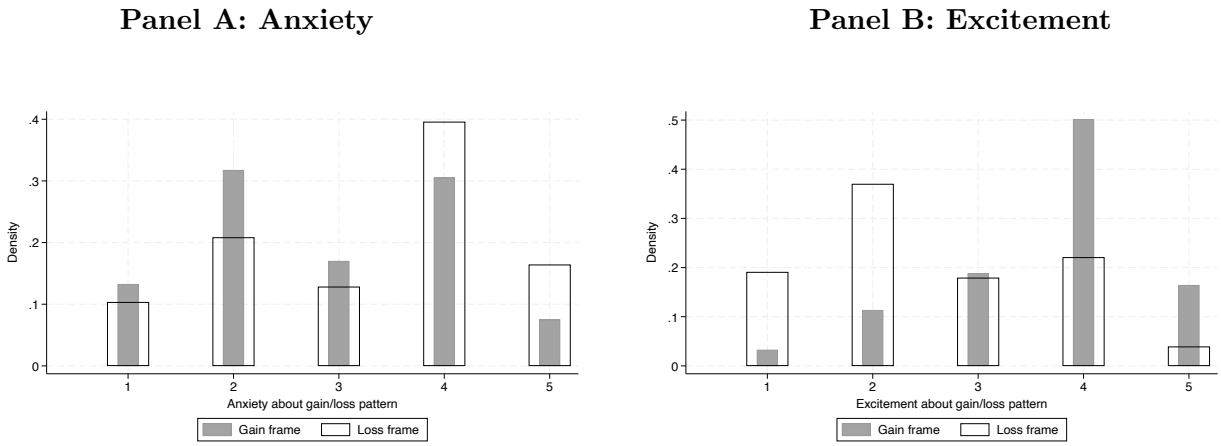


Figure A.2: Histogram of agreement with the statement "During the last set of trials, I felt anxious about whether or not the pattern would be a loss [gain] pattern (the word "anxious" refers to an uneasy, uncomfortable or unpleasant feeling)." (Panel a), or "During the last set of trials, I felt excited about whether or not the pattern would be a loss [gain] pattern (the word "excited" refers to a pleasant or even exhilarating feeling)."" (Panel b) in the Loss Frame [Gain Frame] treatment. Answers are measured on a 5-point Likert scale, split by loss size. Each report in an 8-trial block counts as one observation.

Table A.13: Accuracy in the gain and loss frame treatment.

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy
Loss/No-gain	-4.050	-13.03	-0.848	1.842
	(1.133)	(1.517)	(2.677)	(2.080)
Gain Frame		-6.667		
		(1.524)		
Gain Frame X Loss/No-gain		17.98		
		(2.148)		
Low Emotions (d)			-0.259	
			(2.019)	
Loss/No-gain X Low Emotions (d)			9.580	
			(3.197)	
Risk Averse (d)				0.793
				(1.923)
Loss/No-gain X Risk Averse (d)				6.930
				(3.016)
Constant	63.01	66.36	59.85	59.32
	(0.774)	(1.186)	(1.665)	(1.292)
Treatment	Gain and loss	Gain and loss	Gain only	Gain only
Observations	19200	19200	9600	9600
R^2	0.002	0.011	0.007	0.006

OLS regressions of accuracy on treatment dummies. Columns 1-2 include data from both Gain and Loss Frame domain, columns 3-4 from the Gain Frame only. "Loss/No-gain" is a dummy that equals 1 if the pattern is associated with a Loss (Loss Frame) or the absence of a gain (Gain Frame). "Gain frame" is a dummy that equals 1 in Gain Frame. "Risk Averse" is a dummy that is 1 if a subject self-reports being risk-averse, i.e. scores lower than 4 on the question "How would you evaluate your own attitude towards risk: Are you rather a risk-taking or risk-averse person (trying to avoid risks)? 1: Very risk-averse, 7: Very risk-seeking". "Low Emotions" is a dummy that equals 1 if a subject scores lower than 7 (an approximate median split) on an index that sums the average reported anxiety and excitement in the experiment. Standard errors in parentheses clustered by individual.

F Dynamics

Table A.14: The dynamics of wishful thinking

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy	(7) Accuracy	(8) Accuracy
Loss pattern	-6.519	-17.22	-2.441	-7.374	-5.592	-17.02	-3.590	-6.375
	(2.089)	(2.066)	(1.003)	(1.856)	(1.672)	(1.968)	(0.784)	(1.661)
Trial	0.0411	0.0464	0.0164	0.0240				
	(0.0138)	(0.0299)	(0.0158)	(0.0267)				
Loss pattern x Trial	0.0260	0.0159	-0.0477	-0.0250				
	(0.0156)	(0.0458)	(0.0233)	(0.0388)				
Second half					5.190	2.946	0.0352	2.897
					(1.812)	(1.540)	(0.704)	(1.169)
Loss pattern x Second half					3.422	0.731	-1.566	-2.783
					(2.206)	(2.422)	(0.937)	(1.807)
Experiment	1	2	3	4	1	2	3	4
Observations	11520	11396	33507	21114	11520	11396	33507	21114
R^2	0.008	0.035	0.003	0.010	0.008	0.035	0.003	0.011

OLS regression of accuracy on shock/loss and no-shock/no-loss patterns and temporal indices in various experiments. All regressions include participant fixed effects. Standard errors in parenthesis clustered at the participant level.

Table A.15: Accuracy and response to previous losses

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy	(7) Accuracy	(8) Accuracy
Loss/shock pattern	-3.559	-16.83	-4.200	-8.269	-6.375	-17.52	-2.219	-7.202
	(1.243)	(1.569)	(0.795)	(1.082)	(2.157)	(2.089)	(1.015)	(1.886)
Lagged actual loss	-0.462	-1.655	-0.731	1.134	-0.409	-1.679	-0.699	1.124
	(1.695)	(1.406)	(0.821)	(1.083)	(1.696)	(1.407)	(0.821)	(1.084)
Loss pattern x Lagged loss	-0.933	1.727	-1.487	-0.928	-0.950	1.660	-1.570	-0.932
	(2.677)	(2.257)	(1.268)	(1.717)	(2.702)	(2.259)	(1.266)	(1.718)
Trial					0.0411	0.0464	0.0166	0.0237
					(0.0138)	(0.0299)	(0.0158)	(0.0267)
Loss pattern x Trial					0.0260	0.0161	-0.0484	-0.0250
					(0.0156)	(0.0458)	(0.0232)	(0.0388)
Constant	72.42	77.27	78.14	79.42	67.95	75.36	77.46	78.42
	(0.667)	(0.789)	(0.415)	(0.569)	(1.544)	(1.275)	(0.739)	(1.249)
Experiment	1	2	3	4	1	2	3	4
Observations	11520	11396	33507	21114	11520	11396	33507	21114
R^2	0.002	0.035	0.003	0.010	0.008	0.035	0.003	0.010

OLS regression of accuracy on shock/loss and no-shock/no-loss patterns and lagged shocks or losses. All regressions include participant fixed effects. Standard errors in parenthesis clustered at the participant level.

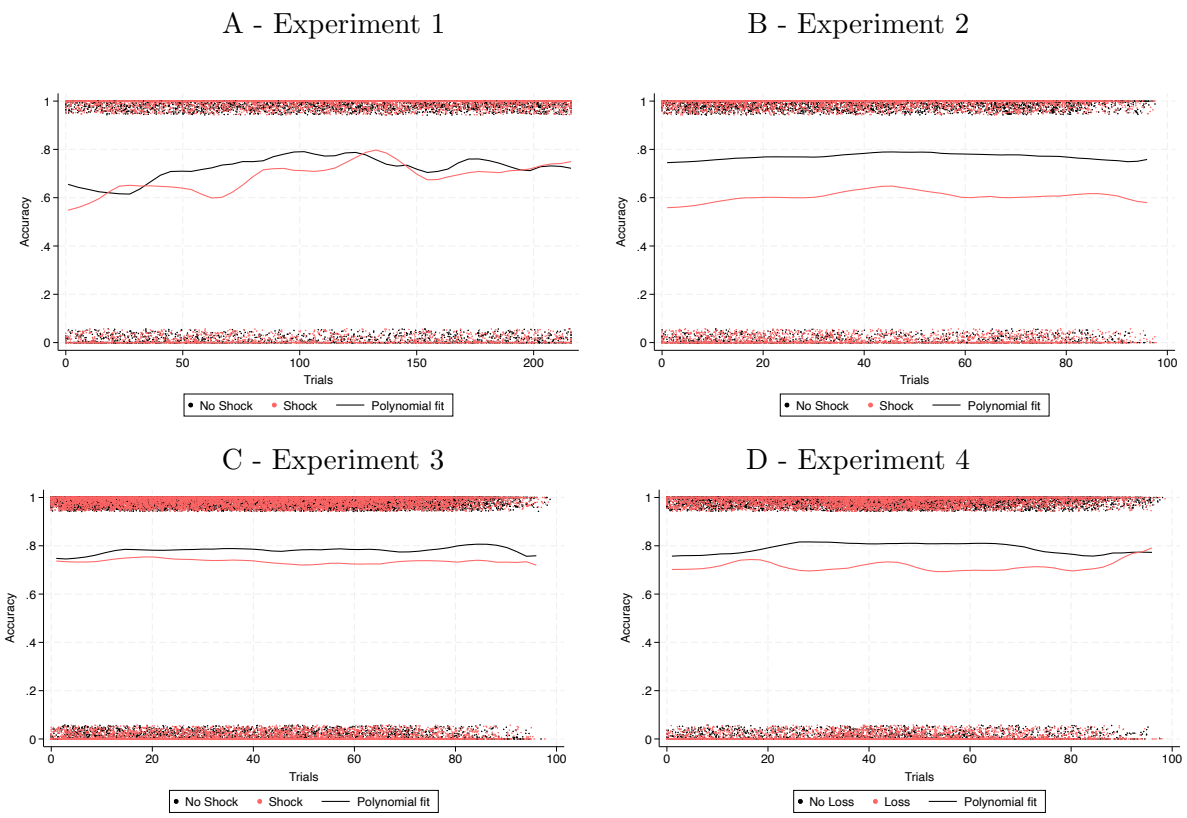


Figure A.3: Dynamics of wishful thinking over time by experiment. x-Axis shows trial numbers. y-Axis is accuracy in single trials (dots - jitter added) and a polynomial fit (line). Note that in Experiment 2, 3, and 4, not all participants completed the same number of trials, as trials stopped when participants reached 5 cumulative (and stochastic) losses from their endowment.

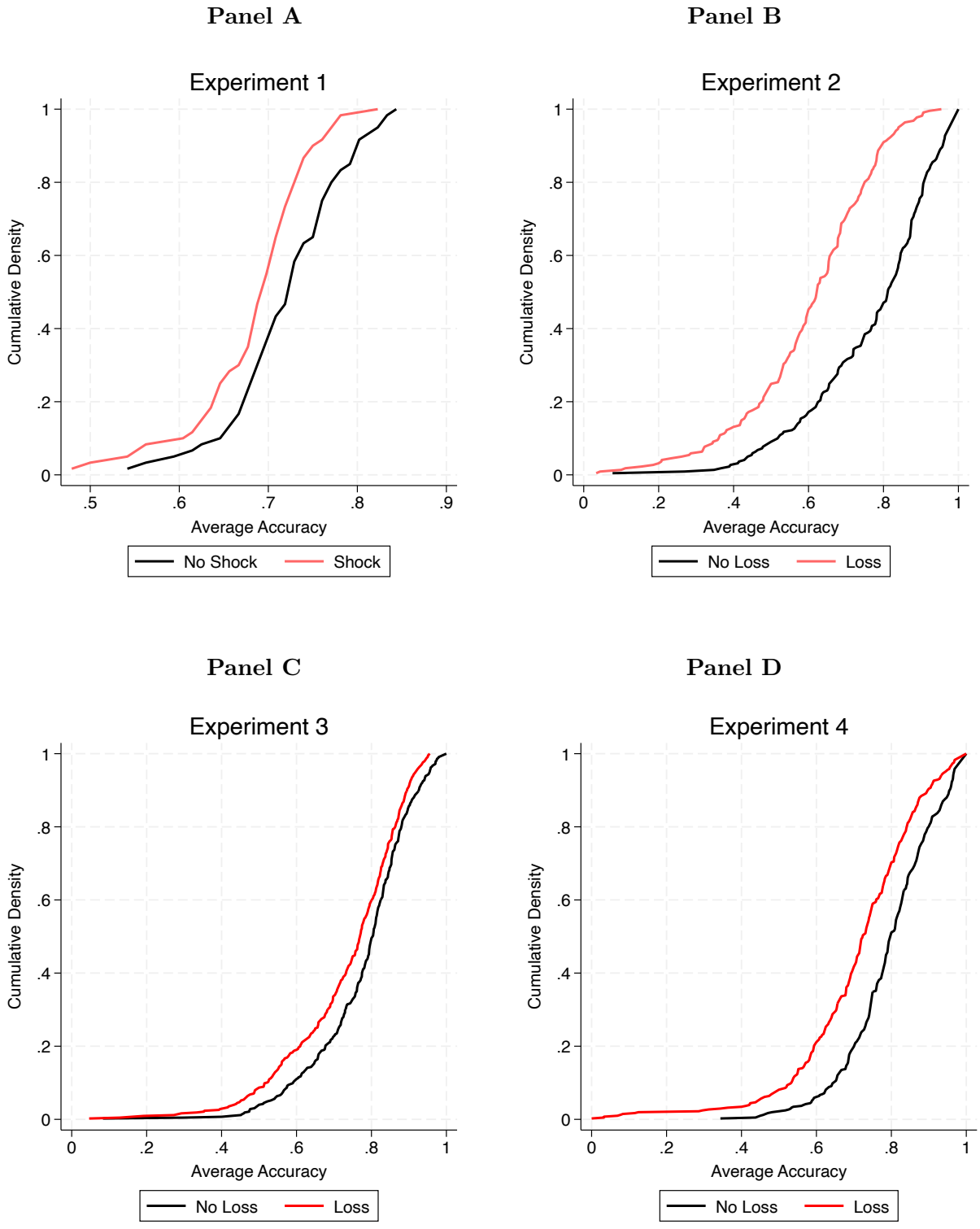


Figure A.4: CDFs of participants' average accuracy in each experiment, split by shock/loss and no-shock/no-loss patterns. Each observation is the average accuracy of all trials of an individual participant in that category.

G Robustness

Table A.16: Accuracy and treatment effect in selected samples

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy	(7) Accuracy
Shock pattern	-7.210	-5.966	-7.005	-6.849	-6.654	-6.390	-6.975
	(0.738)	(0.851)	(0.937)	(0.826)	(0.725)	(0.820)	(0.819)
High accuracy bonus (HAB)	0.695	0.992	0.838	0.976	0.735	0.626	1.306
	(0.508)	(0.632)	(0.685)	(0.595)	(0.534)	(0.575)	(0.593)
Difficult pattern (DP)	-7.996	-8.264	-8.190	-7.903	-8.222	-7.901	-7.987
	(0.347)	(0.416)	(0.453)	(0.383)	(0.350)	(0.393)	(0.385)
Shock pattern x HAB	0.745	0.716	1.418	0.338	1.029	1.034	0.0338
	(0.782)	(0.967)	(1.067)	(0.934)	(0.825)	(0.872)	(0.906)
Shock pattern x DP	-1.328	-1.530	-1.457	-1.656	-1.423	-1.318	-1.188
	(0.481)	(0.571)	(0.633)	(0.533)	(0.474)	(0.523)	(0.533)
Constant	82.21	77.23	77.12	76.31	82.60	81.53	75.80
	(0.740)	(1.227)	(1.328)	(1.109)	(0.665)	(0.845)	(1.100)
Restrictions	None	Difficult Instructions < 4 out of 7	Hard to recall conditions < 5 out of 7	Total no. of mistakes on control q. < 3	Average accuracy > 60 percent	Trust in experimenters > 2 out of 5	Answer does not cause shock
ID clustered S.E.	✓	✓	✓	✓	✓	✓	✓
Experiment fixed effects	✓	✓	✓	✓	✓	✓	✓
Observations	12397	8388	7388	9710	11204	9663	9717
R^2	0.134	0.138	0.135	0.135	0.149	0.130	0.132

OLS regressions of accuracy on treatments across experiments. An observation is the average accuracy per treatment and individual. All regressions include experiment fixed effects and standard errors clustered at the individual level. "Shock pattern" is a dummy representing the pattern is associated with a shock (Experiment 1) or loss (Experiments 2-4). "High accuracy bonus" is a dummy that represents a high accuracy bonus, while "Difficult pattern" is a categorical variable that measures the difficulty of the perceptual task, (see 1 for exact specification by experiment). Column 2 excludes participants with one of the three highest scores on the question "How difficult did you find it to follow the instructions of this experiment?" measured on a 7-point Likert scale. Column 3 excludes participants with one of the three highest scores on the question ""How difficult did you find it to keep in mind information about the potential losses and bonuses associated with trials" measured on a 7-point Likert scale. Column 4 excludes participants who wrongly answered more than two control questions at the beginning of the experiment. Column 5 excludes participants whose average accuracy in the experiment is below 60 percent. Column 6 excludes participants with the three lowest agreement scores with the statement "During the experiment, I never thought that I was deceived by the experimenters about my possible gains or losses" measured on a 5 point Likert scale. Column 7 excludes participants who wrongly answered a multiple choice question about the determinants of the shock. Data in column 2, 3, 4 and 7 exclude Experiment 1, where the relevant measure was not collected.

Table A.17: Accuracy and treatment effect in selected samples, panel regressions

	(1) Accuracy	(2) Accuracy	(3) Accuracy	(4) Accuracy	(5) Accuracy	(6) Accuracy	(7) Accuracy
Shock pattern	-5.989	-5.009	-5.799	-5.799	-5.229	-5.025	-6.202
	(0.679)	(0.785)	(0.861)	(0.757)	(0.631)	(0.748)	(0.758)
High accuracy bonus (HAB)	0.561	0.880	0.678	0.755	0.738	0.629	0.919
	(0.413)	(0.510)	(0.546)	(0.477)	(0.434)	(0.469)	(0.479)
Difficult pattern (DP)	-9.412	-9.815	-9.812	-8.690	-8.929	-9.406	-9.517
	(0.345)	(0.438)	(0.476)	(0.384)	(0.321)	(0.395)	(0.403)
Shock pattern x HAB	0.726	0.911	1.512	0.376	0.811	0.709	0.326
	(0.599)	(0.729)	(0.779)	(0.705)	(0.625)	(0.663)	(0.701)
Shock pattern x DP	-1.842	-2.173	-2.186	-2.314	-1.946	-1.827	-1.778
	(0.435)	(0.531)	(0.591)	(0.499)	(0.422)	(0.475)	(0.502)
Constant	85.14	86.77	87.25	86.17	87.85	85.06	86.55
	(0.424)	(0.510)	(0.549)	(0.553)	(0.421)	(0.481)	(0.483)
Restrictions	None	Difficult Instructions < 4 out of 7	Hard to recall conditions < 4 out of 7	Total no. of mistakes on control q. < 3	Average accuracy > 60 percent	Trust in experimenters > 2 out of 5	Answer does not cause shock
ID clustered S.E.	✓	✓	✓	✓	✓	✓	✓
Individual fixed effects	✓	✓	✓	✓	✓	✓	✓
Observations	77537	47486	42017	55054	68774	61235	53357
R^2	0.040	0.047	0.048			0.039	0.046

Linear regressions of accuracy on treatments across experiments. We use a panel data structure where each observation is a single trial, and regressions include individual fixed effects and standard errors clustered at the individual level. "Shock pattern" is a dummy if the pattern is associated with a shock (Experiment 1) or loss (Experiments 2-4). "High accuracy bonus" is a dummy that represents a high accuracy bonus, while "Difficult pattern" is a categorical variable that measures the difficulty of the perceptual task, (see 1 for exact specification by experiment). Column 2 excludes participants with one of the three highest scores on the question "How difficult did you find it to follow the instructions of this experiment?" measured on a 7-point Likert scale. Column 3 excludes participants with one of the three highest scores on the question ""How difficult did you find it to keep in mind information about the potential losses and bonuses associated with trials" measured on a 7-point Likert scale. Column 4 excludes participants who wrongly answered more than two control questions at the beginning of the experiment. Column 5 excludes participants whose average accuracy in the experiment is below 60 percent. Column 6 excludes participants with the three lowest agreement scores with the statement "During the experiment, I never thought that I was deceived by the experimenters about my possible gains or losses" measured on a 5 point Likert scale. Column 7 excludes participants who wrongly answered a multiple choice question about the determinants of the shock. Data in column 2, 3, 4 and 7 exclude Experiment 1, where the relevant measure was not collected.

Table A.18: Comparison of accuracy of neutral, loss and no-loss patterns

	(1) Accuracy	(2) Accuracy
Loss pattern	-0.110	-0.0449
	(-8.83)	(-5.43)
No-loss pattern	0.0586	0.0398
	(6.52)	(6.35)
Constant	0.712	0.757
	(70.05)	(145.67)
	Experiment 2	Experiment 4
Observations	1105	1221
R^2	0.140	0.065

OLS regression of accuracy on neutral, loss and no-loss patterns in Experiment 2 and 4. Baseline are neutral patterns where no shock was administered present. Each observation is average accuracy per treatment and individual. Standard errors clustered at the participant level in parentheses.

B Heterogeneity

Figure B.1 depicts the histograms of individual-level wishful thinking in experiments 1 through 4. Table B.1 reports half-split correlations. For this exercise we split trials into odd and even numbered trials, trials with easy and trials with difficult patterns, trials in the first half and trials in the second half of the experiment and, for Experiment 2, trials with high stakes and trials with low stakes. Calculating such half-split correlations is common in psychology, where they are used to assess the reliability of individual measures derived from cognitive tasks (for example, see Pronk et al. 2021).[29]

Columns 1 through 3 of Table B.1 report the half-split correlations of wishful thinking.[30] Correlations are around 0.5, with some fluctuations depending on how we split the data, indicating that heterogeneity in wishful thinking reflects individual differences. Moreover, our measure of wishful thinking is only slightly less reliable or stable than participants' skill in the pattern recognition tasks, as shown by the half split correlations in accuracy that we report in columns 4 through 6 of Table B.1. To further show that our results are not driven by a few outliers, Figure B.2 shows the scatterplots pertaining to the odd-even trial splits in Table B.1.

[29] In the same vein, our results here also help us assess how reliably our experimental design identifies wishful thinking.

[30] We exclude Experiment 1 because there we recalibrated both the strength of the shock and the difficulty of the patterns during the experiment. This confounds the half-split correlations of wishful thinking and accuracy.

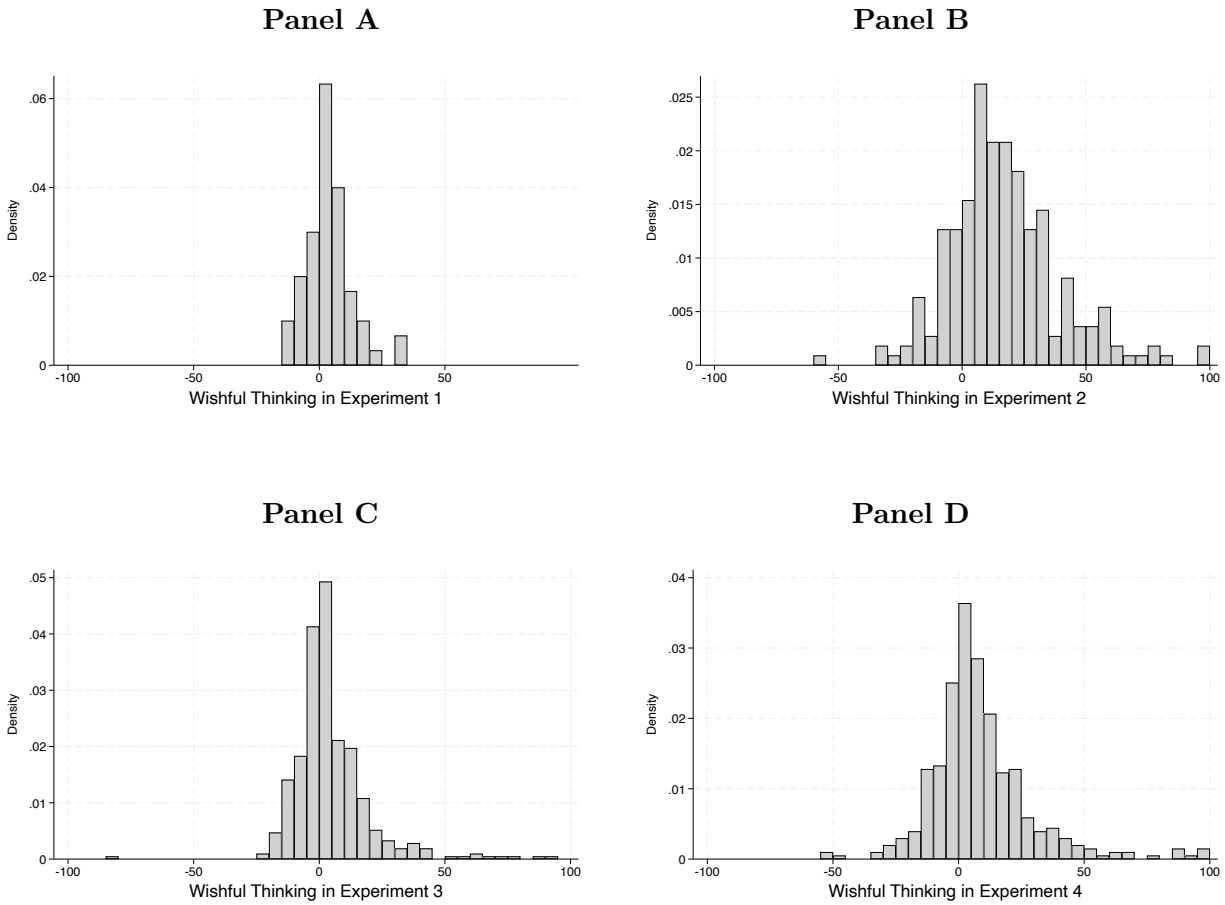


Figure B.1: Histograms of participants' wishful thinking in each experiment. Wishful thinking is defined as an individual's accuracy for shock patterns minus their accuracy for no shock patterns.

Table B.1: Half-split correlations

	Wishful thinking			Accuracy		
	(1)	(2)	(3)	(4)	(5)	(6)
X/Y	Exp. 2	Exp. 3	Exp. 4	Exp. 2	Exp. 3	Exp. 4
Odd/even trials	0.641	0.461	0.570	0.592	0.730	0.476
Difficult/easy patterns	0.573	0.526	0.457	0.575	0.497	0.563
First/second half	0.441	0.435	0.350	0.663	0.568	0.478
Low/high losses	0.460	-	-	0.589	-	-

Correlations between individual participants' wishful thinking or accuracy as measured in X and Y trials. X and Y correspond to odd and even, difficult and easy, first and second half, and low and high loss trials respectively.

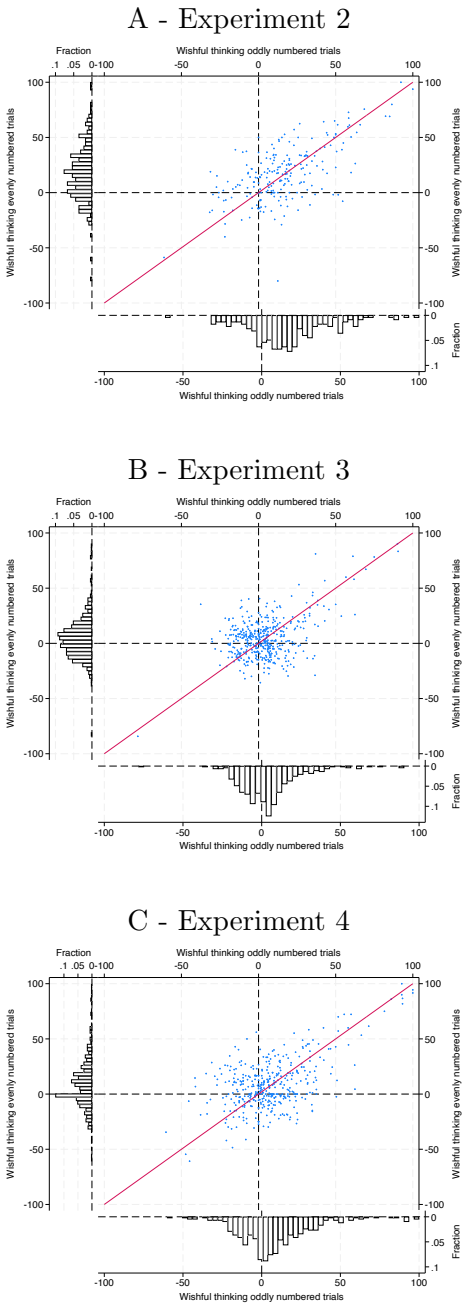


Figure B.2: Scatterplot of participants' wishful thinking in odd and even trials in each experiment. Wishful thinking is defined as an individual's accuracy for shock patterns minus their accuracy for no-shock patterns. Each plot includes the 45 degree line as well as projected histograms for odd and even trials.

C Replication Experiment 1

Before conducting Experiment 1, we ran an experiment with an almost identical design, which was also preregistered on aspredicted.org (preregistration is here). There were some small differences. First, the experiment also featured some neutral trial blocks in which subjects did not face the threat of a shock. Second, while we used the same visual patterns for the task, subjects had to indicate whether they were vertically or horizontally oriented (rather than choosing the closest diagonal), and there were four difficulty levels. Third, incentives were constant across the experiment. Finally, the experimental code exhibited a small bug which meant that the ambiguity levels were not equally calibrated across the Shock and No-shock condition. While we are able to control for the ambiguity level (see below), the imperfect randomization ultimately caused us to rerun this experiment, resulting in Experiment 1. For our purposes, two aspects of the precursor experiment are of interest. First, we investigate whether our main treatment effect obtains also in this study. The experiment replicates our main results with very similar effect sizes (μ) between shock and no-shock patterns (Accuracy: $\mu = 0.047$, $s.d. = 0.103$, $p = 0.0024$; Belief: $\mu = 0.031$, $s.d. = 0.069$, $p = 0.0022$). Table C.1 shows the result of an OLS regression of Accuracy and Belief, averaged by subject and condition, on treatment dummies. Because pattern difficulty was not well balanced between the Shock and No-Shock conditions, we control for difficulty in Table C.1. This shows similar results, with a highly significant effect of the Shock condition which is 0.037 for Accuracy and 0.029 for Belief, closely mirroring the magnitudes in our main experiment.

Second, the presence of "Neutral" trials without the threat of shocks allows us to investigate the "noise explanation" elaborated in Section VII.C, where we inspect how the presence of a shock-threat affects Accuracy and Beliefs. We find that the Accuracy and Belief in the Neutral patterns are substantially worse than in the No-Shock patterns. This shows that the presence of a shock does not necessarily reduce performance, reinforcing our confidence that wishful thinking is driving our main result.

	(1) Accuracy	(2) Belief
Shock Pattern	-3.734	-2.854
	(1.521)	(1.034)
Neutral Pattern	-3.392	-3.714
	(1.140)	(0.736)
Difficulty	-3.490	-2.553
	(0.497)	(0.411)
Constant	90.38	80.98
	(1.323)	(1.268)
Observations	600	600
R^2	0.098	0.087

Table C.1: OLS regressions of Accuracy and Belief on the experimental conditions. Standard errors in parentheses clustered at subject level.

D Theory extensions

A Anticipatory utility from accuracy incentives

In the following, we suppose that the agent also obtains utility from her anticipation of incentives for accuracy. Her utility then takes the following form.

$$\begin{aligned} U = \quad & \frac{1}{2}\left(1 + 2p\hat{p} - \hat{p}^2\right) M + \sigma_m \frac{1}{2}\left(1 + \hat{p}^2\right) M \\ & - (r_z p + (1 - r_z)(1 - p))qZ - \sigma_z(r_z \hat{p} + (1 - r_z)(1 - \hat{p}))qZ \\ & - \lambda(s)(p - \hat{p})^2, \end{aligned}$$

where σ_m captures the agent's propensity to savor future payoffs from the accuracy incentives and is distinct from the anxiety of being shocked or losing money, which is still parameterized by σ_z. Note that the agent's anticipatory utility from expecting future accuracy payoffs depends only on her chosen belief $\hat{p}$ and not on the undistorted or true probability of the pattern being right-tilted p.

Maximizing the agent's utility then yields the following optimal beliefs

$$\hat{p}^* = \frac{M + 2\lambda(s)}{(1 - \sigma_m)M + 2\lambda(s)} p(s, r_\theta) - \frac{\sigma_z(2r_z - 1)qZ}{(1 - \sigma_m)M + 2\lambda(s)},$$

and wishful thinking is given by

$$W := \hat{p}^*(r_z = 0) - \hat{p}^*(r_z = 1) = \frac{2\sigma_z qZ}{(1 - \sigma_m)M + 2\lambda(s)}.$$

We see that an anticipatory utility motive stemming from the savoring of accuracy incentives does not change qualitative predictions regarding the drivers of anxiety-induced wishful thinking. In particular, the comparative statics of wishful thinking in M, s and Z have the same sign they had in the main model. However, wishful thinking is now increasing in the savoring parameter σ_m because the higher σ_m, the more the agent cares about her perceived rather than her actual receipt of accuracy incentives, and the former is not decreasing in the amount of belief distortion.

B Defensive pessimism and bracing

To capture defensive pessimism or bracing we suppose that the agent maximizes the following utility function.[31]

$$\begin{aligned} U = \quad & \frac{1}{2}\left(1 + 2p\hat{p} - \hat{p}^2\right) M \\ & - (r_z p + (1 - r_z)(1 - p))q(Z - b\hat{p}Z) - \sigma_z(r_z\hat{p} + (1 - r_z)(1 - \hat{p}))q(Z - b\hat{p}Z) \\ & - \lambda(s)(p - \hat{p})^2. \end{aligned}$$

Here, parameter b captures the benefit of bracing or the extent to which defensive pessimism can soften the blow of a shock or loss. If $b = 0$, pessimistic beliefs do not lessen the impact of a shock. In the other extreme, $b = 1$, the agent can fully negate the shock's impact by being maximally pessimistic. The agent's optimal beliefs are then given by

$$\begin{aligned} \hat{p}^*(r_z = 1) &= \frac{M + 2\lambda(s) + bZ}{M + 2\lambda(s) - 2\sigma_z qbZ}p - \frac{\sigma_z qZ}{M + 2\lambda(s) - 2\sigma_z qbZ} \\ \hat{p}^*(r_z = 0) &= \frac{M + 2\lambda(s) - bZ}{M + 2\lambda(s) + 2\sigma_z qbZ}p + \frac{\sigma_z qZ}{M + 2\lambda(s) + 2\sigma_z qbZ}. \end{aligned}$$

We can check that $\frac{d\hat{p}^*(r_z=1)}{db} > 0$ and $\frac{d\hat{p}^*(r_z=0)}{db} < 0$. So the bracing motive decreases apparent wishful thinking, which is defined as $W := \hat{p}^*(r_z = 0) - \hat{p}^*(r_z = 1)$. Furthermore, it is easy to verify that there is wishful thinking, i.e. $\hat{p}^*(r_z = 1) < p$ and $\hat{p}^*(r_z = 0) > p$, only if the following inequality holds

$$b < \frac{\sigma_z q}{1 + 2\sigma_z q}.$$

This inequality is satisfied for small b and large σ_z. Therefore, if we detect wishful thinking in our average participant, then we provide evidence not just for wishful thinking but for the fact that the anticipatory anxiety motive dominates the defensive pessimism or bracing motive.

C The correlation between anxiety and wishful thinking

The agent's utility function contains a term that captures her experienced anxiety conditional on her (optimal) belief and the tilt of the shock pattern. Since the shock pattern is right-tilted 50

[31] We take "bracing" to mean any action or investment that can reduce the impact (physical or otherwise) of negative news or events. Defensive pessimism is a more specific concept and form of bracing. It is a cognitive strategy that allows one to deal with the psychological impact of negative events by holding negative expectations (Norem and Cantor, 1986).

percent of the time and left-tilted 50 percent of the time, average experienced anxiety is given by

$$A = \frac{1}{2}\sigma_z \hat{p}^*_{r_z=1} qZ + \frac{1}{2}\sigma_z (1 - \hat{p}^*_{r_z=0}) qZ$$

Substituting into this term her respective optimal beliefs $\hat{p}^*_{r_z=1}$ and $\hat{p}^*_{r_z=0}$ from the main model yields

$$A = \frac{1}{2}\sigma_z qZ - \frac{\sigma_z^2 q^2 Z^2}{M + 2\lambda}.$$

Next, we turn to the comparative statics of A in λ and σ_z. Comparing these with the comparative statics of W in λ and σ_z, will allow us to see how heterogeneities in λ and σ_z map into correlations between experienced anxiety and wishful thinking.

It is easy to see that $\frac{dA}{d\lambda} > 0$, so that experienced anxiety is increasing in the cognitive costs of self-deception. How A varies with σ_z is more subtle. We can show that $\frac{dA}{d\sigma_z} > 0$ if and only if

$$\sigma_z < \frac{4q^2 Z^2}{M + 2\lambda}, \tag{C.1}$$

and $\frac{dA}{d\sigma_z} \leq 0$ otherwise. So experienced anxiety A is increasing in innate anxiety σ_z for small levels of innate anxiety and decreasing for higher levels. To see why experienced anxiety must eventually decrease in innate anxiety note that for very high levels of innate anxiety the agent will engage in sufficient wishful thinking to put zero subjective probability on the negative outcome, leaving her with no experienced anxiety.

We can use our experimental results to get a sense of where in the parameter space participants are likely to be located. To this end, note that inequality C.1 can be rewritten to state that $\frac{dA}{d\sigma_z} > 0$ if and only if

$$W < \frac{1}{2}$$

Clearly, average wishful thinking is well below 50 percentage points in all experiments. Going forward we therefore assume that inequality C.1 is satisfied.

Putting together the comparative statics on σ_z and λ, we now consider two scenarios.

Secenario A (heterogeneity in λ). Suppose there are two groups of participants (group 1 and group 2) that differ only in λ, their cognitive costs of self-deception. In particular, suppose that $\lambda_1 < \lambda_2$, where the subscript is the group label. Based on how W and A vary with λ, it will then be the case that $W_1 > W_2$ and $A_1 < A_2$. Therefore, in this scenario, wishful thinking and

experienced anxiety are negatively correlated, as those with low cost of self-deception have higher wishful thinking and hence lower experienced anxiety.

Secenario B (heterogeneity in σ_z). Now suppose there are two groups of participants (group 1 and group 2) that differ only in σ_z, their innate anxiety. In particular, suppose that $\sigma_{z1} < \sigma_{z2}$, where the subscript is the group label. Based on how W and A vary with σ_z, it will then be the case that $W_1 < W_2$ and $A_1 < A_2$. Therefore, in this scenario, wishful thinking and experienced anxiety are positively correlated, as high innate anxiety leads to the higher experienced anxiety, despite a partial offset by higher wishful thinking.

Considering these two scenarios allows us to conclude that, according to the model, whether measures of experienced anxiety and wishful thinking are positively, negatively, or not correlated is ambiguous. More specifically, it will depend on whether there is a dominant heterogeneity and whether this heterogeneity is in participants' ability to self-deceive σ_z (negative correlation) or in their innate anxiety λ (positive correlation).

D Optimal beliefs with a hard cognitive constraint

In this section we describe a model that better captures the statistical relationships we see in the data. We will add two elements to the model. First, self-deception is constrained to some maximum amount or hard cognitive constraint. One interpretation, closest to our original model and developed below in more detail, is that this is a binding constraint on an optimizing agent who chooses optimal beliefs. A second interpretation is that self-deception is an "automatic" or "system 1" process, where a certain amount of self-deception occurs automatically without an agent's cognitive influence. The second new element is that the constraint on self-deception depends on the signal strength, which is determined by the agent's investment in signal precision or information-gathering.

Thus, we suppose that the agent first invests in signal precision at time $t = 0$ and then distorts her mental representation of the signal at time $t = 1$. To solve the agent's problem, we first look at $t = 1$.

D.1 Belief choice conditional on the signal at $t = 1$

Consider the following maximization problem.

$$
\begin{aligned}
Max \quad U_1 = \quad & \frac{1}{2}\left(1 + 2p\hat{p} - \hat{p}^2\right) M \\
& - (r_z p + (1 - r_z)(1 - p))qZ - \sigma_z(r_z \hat{p} + (1 - r_z)(1 - \hat{p}))qZ \\
\text{such that} \quad & |p - \hat{p}| \leq \lambda(s)
\end{aligned}
$$

So self-deception is cognitively costless up to a point and then becomes impossible. In line with our results, we further assume that the maximum distance between true and distorted beliefs is decreasing in s, the signal precision, i.e. $\lambda'(s) < 0$.

Solving this maximization problem yields the following optimal beliefs.

$$
\hat{p}^* = \begin{cases} \max(p - \frac{\sigma_z qZ}{M}, p - \lambda(s)) & \text{if} \quad r_z = 1 \\ \min(p + \frac{\sigma_z qZ}{M}, p + \lambda(s)) & \text{if} \quad r_z = 0 \end{cases}
$$

Our results in Experiments 1 to 3 indicate that ex-post signal distortion responds to s, but not to M or Z. In other words, the cognitive constraint is binding for all values of M and Z and optimal beliefs are given by $p - \lambda(s)$ for right-tilted patterns and by $p + \lambda(s)$ for left-tilted patterns. In this case, wishful thinking is given by $W = 2\lambda(s)$, which is decreasing in signal precision s.

In a next step we look the an agent's investment in signal precision.

D.2 Investment in signal precision at $t = 0$

The agent decides on her investment in signal precision knowing whether shocks are associated with left or with right-tilted patterns and anticipating the effect of signal precision on her payments from the BDM mechanism and her ability to self-deceive. At the point of deciding on the cognitive effort she spends on identifying the pattern, she does not know the actual tilt of the pattern and merely has a 50:50 prior over whether the pattern is left- or right-tilted. Her choice of signal precision s maximizes the following function.

$$
\begin{aligned}
U_0 = \quad & \frac{1}{2}\left(\hat{p}_{r_t=1} M + (1 - \hat{p}_{r_t=1})\frac{1 + \hat{p}_{r_t=1}}{2} M\right) + \frac{1}{2}\left((1 - \hat{p}_{r_t=0})\frac{1 + \hat{p}_{r_t=0}}{2} M\right) \\
& - \frac{1}{2}\sigma_Z(r_z \hat{p}_{r_t=1} + (1 - r_z)(1 - \hat{p}_{r_t=1}))qZ - \frac{1}{2}\sigma_Z(r_z \hat{p}_{r_t=0} + (1 - r_z)(1 - \hat{p}_{r_t=0}))qZ \\
& - \frac{1}{2} qZ - c(s),
\end{aligned} \tag{C.2}
$$

where $c(s)$ is the cognitive cost associated with generating more precise representations of the signal. Moreover, $\hat{p}_{r_t=\{0,1\}}$ are the agent's optimal $t = 1$ beliefs, conditional on the true pattern, that depend on $p_{r_t=\{0,1\}}$, her undistorted belief, which in turn depends on $s \in [0.5, 1]$.

For simplicity, we assume that undistorted beliefs are $p_{r_t=1} = s$ and $p_{r_t=0} = 1 - s$. So as s increases, the agent becomes more accurate in identifying both right- and left-tilted patterns. Furthermore, we assume that $\lambda(s) = \epsilon(1 - s)$ and that $c(s) = cs^2$.

We consider the case of $r_z = 1$ so that $\hat{p}_{r_t} = p_{r_t} - \lambda(s)$. Then, substituting the expressions for $p_{r_t=\{0,1\}}$, $\lambda(s)$ and $c(s) = cs^2$ into C.2 and simplifying yields

$$\begin{aligned} U_0 = \quad & \frac{1}{2}M\left(\frac{1}{2} + 2s - s^2 - \frac{1}{2}\epsilon^2(1-s)\right) \\ & - \frac{1}{2}\sigma_Z(1 - \epsilon + \epsilon s)qZ \\ & - \frac{1}{2}qZ - cs^2 \end{aligned}$$

The s that maximizes this ex ante utility is given by

$$s^* = \frac{\frac{1}{2}M(2+\epsilon^2) - \frac{1}{2}\sigma_z \epsilon q Z}{\frac{1}{2}M(2+\epsilon^2) + 2c} \tag{C.3}$$

Taking the first derivative yields that $\frac{ds^*}{dM} > 0$. So an increase in accuracy incentives increases signal precision. Then, because $\lambda'(s) < 0$ and wishful thinking is decreasing in λ, we have that $\frac{dW}{dM} < 0$.

Note that this chain of reasoning depends on the agent being *able* to increase her signal precision, i.e. $\frac{ds^*}{dM} > 0$. Across our experiment, only dot counters in experiment 4 were able to increase their accuracy in pattern recognition (a proxy for signal precision) in response to an increase in incentives.

Hypothesis C.1 (Incentives) *If signal precision and hence, accuracy, is increasing in accuracy incentives $\frac{ds^*}{dM} > 0$, then wishful thinking is decreasing in accuracy incentives, i.e. $\frac{dW}{dM} < 0$.*

Naivite and sophistication. Here we assumed that the agent is sophisticated about the effect of her investment in s on her ability to self-deceive. An agent who is naive about the link between signal precision and her subsequent ability to self-deceive expects that $\lambda'(s) = 0$. The naive agent's optimal signal precision is then given by

$$s_n^* = \frac{M}{M + 2c} \tag{C.4}$$

For the naive it will therefore also be the case that $\frac{ds_n^*}{dM} > 0$. Because, in reality, $\lambda(s) > 0$ and $\lambda'(s) < 0$, a naive's wishful thinking will then also be decreasing in M. This implies that the result that wishful thinking is decreasing in M does not help us distinguish between naives and sophisticates. However, note that s^* but not s_n^* are increasing in Z, the size of the loss or shock. A positive effect of Z on wishful thinking is therefore suggestive of sophistication.

E IRB and Preregistration

UNIVERSITY OF AMSTERDAM

Ethics Committee Economics and Business (EBEC)
University of Amsterdam

Amsterdam Business School
Plantage Muidergracht 12
1012 TV Amsterdam
The Netherlands
T +31 20 525 7384
www.abs.uva.nl

To: van der Weele

Date	Our reference	
June 12, 2017	EC 20170510120541	
Contact	Telephone	E-Mail
Sophia de Jong	(31)20-5255311	secbs-abs@uva.nl

Subject
EBEC approval

Dear Joel van der Weele,

The Economics & Business Ethics Committee (University of Amsterdam) received your request nr 20170510120541 to approve your project "Anticipatory utility and probabilistic confidence".

We evaluated your proposed research in terms of potential impact of the research on the participants, the level and types of information and explanation provided to the participants at various stages of the research process, the team's expertise in conducting the proposed analyses and particularly in terms of restricted access to the data to guarantee optimal levels of anonymity to the participants.

The Ethics Committee approves of your request.

Best regards,
On behalf of the Ethics Committee Economics and Business,

Prof. Dr. J.H. Sonnemans
Chairman of the Committee

UNIVERSITY OF AMSTERDAM

Ethics Committee Economics and Business (EBEC)
University of Amsterdam

Amsterdam Business School
Plantage Muidergracht 12
1012 TV Amsterdam
The Netherlands
T +31 20 525 7384
www.abs.uva.nl

To: van der Weele

Date	Our reference	
February 02, 2021	EC 20210202020244	
Contact	Telephone	E-Mail
Sophia de Jong	(31)20-5255311	secbs-abs@uva.nl

Subject
EBEC approval

Dear Joel van der Weele,

The Economics & Business Ethics Committee (University of Amsterdam) received your request nr 20210202020244 to approve your project "Anticipatory anxiety from monetary losses and wishful thinking ".

We evaluated your proposed research in terms of potential impact of the research on the participants, the level and types of information and explanation provided to the participants at various stages of the research process, the team's expertise in conducting the proposed analyses and particularly in terms of restricted access to the data to guarantee optimal levels of anonymity to the participants.

The Ethics Committee approves of your request.

The information as filled in the form, can be found at
https://www.creedexperiment.nl/EBEC/showprojectAVG.php?nummer=20210202020244

Best regards,
On behalf of the Ethics Committee Economics and Business,

Prof. Dr. J.H. Sonnemans
Chairman of the Committee

Wishful thinking and anxiety in the laboratory (#15709)

Created: 10/29/2018 02:38 PM (PT)
Shared: 01/15/2019 11:37 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
1) If an uncertain outcome is associated with painful consequences, does that lead people to engage in wishful thinking by underestimating the probability of that outcome?
2) Can higher incentives for accuracy reduce wishful thinking?

3) Describe the key dependent variable(s) specifying how they will be measured.
There are two key dependent variables:
1) Accuracy: this is a binary variable (1:correct; 0:incorrect), indexing at each trial whether the participant correctly identified the displayed pattern.
2) Confidence: this is elicited trial-by-trial, as a number between 50% and 100%: after each choice participants are asked to report the probability that this answer is correct (50% is chance level, and 100% is certainty). On the basis of this we can construct a “belief” measure, which measures on a scale from 0 to 100% the belief of the subject about the orientation of the true pattern.

4) How many and which conditions will participants be assigned to?
In total, a participant sees 216 Gabor patches, and has to recognize whether these patterns are tilted to the right or left. After deciding between these two answers, the participant indicates his confidence in this decision in percentages. The participants are incentivized monetarily for providing accurate answers by a matching probability. At the start of the experiment, each participant is connected to an electric stimulation device that is personally calibrated to deliver mild but unpleasant shocks. Participants will receive an electric shock with a probability of 1/3 if the true answer is either right or left (depending on the condition).

This is a 2x2 design with 4 conditions. Each participant will participate in each condition (a within-subject design), and the pattern recognition tasks are equally divided over all conditions (54 in each condition). The two treatments dimensions are a) the incentives for accuracy (high and low), and b) whether the shock is associated with the left-leaning Gabor patch or the right-leaning Gabor patch.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We are interested how the shock influences the accuracy and confidence of the participants. Specifically, we will test the following directional hypotheses:

1) Wishful thinking 1: Does the accuracy of identifying a given pattern go down if the potential shock is aligned with the true answer (the unpleasant true answer), relative to the case where the potential shock is not aligned with the answer? We use one-sided t-tests to evaluate the differences in the average accuracy and confidence between conditions, where each observation is the average of a subject’s answers in that condition.

2) Wishful thinking 2: Similarly, we will use one-sided t-tests to examine whether the confidence in the true answer decreases if the potential shock is aligned with the true answer.

3) Accuracy incentives: We will test use one-sided t-tests whether accuracy and confidence in the true answer are higher in the condition with high incentives for accuracy.

In addition to t-tests, we will also use multivariate linear regression analysis to test the effect of our treatments (accuracy incentives, shock alignment) as well as their interaction. Finally, we will use linear mixed effect models (with or without individual fixed effects) where we can control for trial characteristics and/or subject characteristics (see below).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We calibrate the difficulty of the task in the beginning of the experiment, so that we expect participants to be accurate 75% of the time on average. The actual accuracy in the experiment may deviate from this, and we will exclude the participant if actual accuracy is outside the [60%-90%] range, as this may indicate that, despite the calibration, the task was either too easy or too hard to detect meaningful differences.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We ran a previous study that we had to discard due to an error in the code but was very similar. On the basis this study, running the same tests as specified above, we calculated that we could achieve more than 80% power with a sample of 60 people, so we will invite 60 participants

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We are likely to do some further exploratory analysis. For instance, we will see if the strength of the effect size differs by the difficulty of the task. We also look whether people who score higher on the trait anxiety, which we measure with a psychological questionnaire, are more affected by the shock, both in their accuracy and their beliefs. To investigate this, we will run mixed models which feature interactions between the shock treatment and the trait anxiety measured by the questionnaire.

Finally, we might explore how the effects of shocks play in typical models of confidence formation (inspired from signal-detection theory): confidence is known to be an increasing function of evidence for correct answers and decreasing for incorrect answers. We can test if the presence of shocks modulate the intercept of the slopes of this model (see e.g. Lebreton, et al. (2017) biorXiv for similar analysis)

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Anticipatory anxiety about monetary losses and wishful thinking 2021 (#57718)

Created: 02/08/2021 03:07 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Our main question is about "wishful thinking": If an ambiguous state-of-the-world is associated with monetary losses, does that lead people to be less likely to correctly identify that state?
Secondary hypotheses: Does wishful thinking increase with…
1)…lower incentives for accuracy?
2)…higher monetary losses?
3)…increased ambiguity of the evidence?

3) Describe the key dependent variable(s) specifying how they will be measured.
A participant sees up to 96 Gabor patches, and has to recognize whether these patterns are tilted to the right or left. The main dependent variable is based on their accuracy on this task in each trial, i.e. whether the participant correctly identified the displayed pattern. Wishful thinking is constructed as the difference in accuracy in recognizing patterns associated with monetary losses and those not associated with such losses.

4) How many and which conditions will participants be assigned to?
Participants are endowed with an amount of money. They can lose part of this endowment on each trial with a probability of 1/3 if the true state is either right or left (depending on the condition) – i.e. monetary losses are not associated with participants' answers, but with the Gabor actual tilt direction. Additionally, participants are rewarded monetarily for providing accurate answers on the perception task.

There are 4 treatment dimensions. Each participant will participate in each condition (a within-subject design), and the pattern recognition tasks are equally divided over all treatments. The treatments dimensions are a) monetary loss is associated with the left-leaning Gabor patch or the right-leaning Gabor patch, b) the size of those losses (zero, low, or high), (c) the incentives for accuracy (high and low), and d) the ambiguity of the pattern (easy versus hard to discriminate).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We are interested how the losses influences the accuracy and confidence of the participants. Specifically, we will test the following directional hypotheses:

1) Wishful thinking (main hypothesis): Does the accuracy of identifying a given pattern go down if the potential loss is aligned with the Gabor patch reflecting the correct answer, relative to the case where the potential shock is not aligned with the answer? We use a t-test to evaluate the differences in the average accuracy and confidence between conditions, where each observation is the average of a subject's answers in that condition.

2) Secondary hypotheses: We will test use t-tests to assess whether wishful thinking is higher in the conditions with (a) low incentives for accuracy, (b) higher ambiguity of the pattern, and (c) higher potential losses.

In addition to t-tests, we will also use multivariate linear regression analysis to test the effect of our treatments (accuracy incentives, loss alignment) as well as their interaction. We will use linear mixed effect models where we can control for trial characteristics and/or subject characteristics. Finally, we will study the effect of incentives for accuracy on accuracy in the task.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Our experiment takes place on prolific. We will exclude participants who fail to answer simple attention checks at the beginning and throughout the experiment.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit 220 subjects to the experiment.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
More exploratory analysis will correlate wishful thinking with questionnaire items related to the self-reported anxiety of subjects about losing money, as

well as other questionnaire responses that include trait anxiety and emotion regulation.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Anticipatory anxiety about monetary losses, wishful thinking, and the effe (#83830)

Created: 12/21/2021 10:06 AM (PT)

---

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
In previous experiments we have found evidence for "wishful thinking". If a state of the world is associated with aversive outcomes, then subjects are less likely to correctly identify that state. Here, we will test whether wishful thinking is motivated by the threat of monetary losses. However, the main directional hypothesis we test is whether wishful thinking decreases with higher incentives for accuracy.

3) Describe the key dependent variable(s) specifying how they will be measured.
A participant sees several short sequences of Gabor patches, and is asked to recognize whether these sequences are more right- or more left-tilted. The main dependent variable is based on their accuracy on this task in each trial, i.e., whether the participant correctly identified the displayed patterns. Wishful thinking is constructed as the difference in accuracy in recognizing patterns not associated with monetary losses and patterns associated with such losses. Our paradigm expands on an earlier experiment, which focused on the accurate recognition of the tilt of a single Gabor patch.

4) How many and which conditions will participants be assigned to?
Participants are endowed with an amount of money. They can lose part of this endowment on each trial with a probability of 1/3 if the true state is either right or left tilted (depending on the condition) – i.e., these monetary losses are not associated with participants' answers, but with the actual tilt of the sequence. In addition, participants are rewarded with a monetary bonus for correctly identifying sequences.

There are 3 treatment dimensions. Each participant will participate in each condition (a within-subject design) and the pattern recognition tasks are equally divided across treatments. The treatment dimensions are a) monetary loss is associated with a left-leaning sequence of Gabor patches or the right-leaning sequence, b) the incentives for accuracy are high or low, and c) the ambiguity of the sequences of patterns is high or low. Our measure for ambiguity derives from a continuous variable of pattern difficulty that we dichotomize. The experimental uses a 2 x 2 x 2 within-subjects design with the factors loss pattern (aligned, not aligned with loss), incentives (high, low) and ambiguity (high, low).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will run a linear regression of correct answers on the following variables
- [loss pattern]: a dummy for whether the pattern is associated with a loss.
- [high incentives]: a dummy for the high incentive condition.
- [interaction]: a dummy for the interaction of the two conditions.

We cluster standard errors at the individual level. We run regressions where each observation is the individual average over trials in that condition, as well as regressions where we use all data points.

We test for
- wishful thinking by testing whether the first coefficient is positive and stat. significant.
- the effect of incentives on overall task performance by testing whether the second coefficient is positive and stat. significant.
- the interaction (our main hypothesis), by testing whether the third coefficient is positive and stat. significant.

For patterns associated with a loss, incentives should raise performance both because of the overall effect and because of a reduction in wishful thinking. To test this joint effect, we conduct an additional t-test assessing whether high accuracy incentives improve performance on loss patterns only.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We conduct our online experiment on Prolific. We will exclude participants who fail to answer simple attention checks at the beginning and throughout the experiment.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
400

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will test whether self-reported concentration responds to incentives, as a manipulation check.
We will test our secondary hypothesis that wishful thinking increases with the ambiguity of the evidence.
Leveraging our within subject identification of wishful thinking, we will also study the heterogeneity of wishful thinking across subjects, correlating it with responses from the exit questionnaire.
We will test the robustness of our results if we exclude participants whose average accuracy is worse than chance.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Anticipatory anxiety, wishful thinking, and the effect of accuracy incentives in a dot-task. (#89876)

Created: 03/04/2022 02:58 AM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
In previous experiments we have found evidence for "wishful thinking". If a state of the world is associated with aversive outcomes, then subjects are less likely to correctly identify that state.
Here, we will test whether wishful thinking is motivated by the threat of monetary losses in a task we have not previously used. However, the main directional hypothesis we test is whether wishful thinking decreases with higher incentives for accuracy.

3) Describe the key dependent variable(s) specifying how they will be measured.
On each trial, a participant sees an array of 100 blue and red dots, and is asked to recognize whether there are more blue or red dots. The main dependent variable is based on their accuracy on this task in each trial, i.e., whether the participant correctly identified the displayed patterns. Wishful thinking is constructed as the difference in accuracy in recognizing patterns not associated with monetary losses and patterns associated with such losses.

Our paradigm expands on an earlier experiment, which focused on the accurate recognition of the tilt of Gabor patches.

4) How many and which conditions will participants be assigned to?
Participants are endowed with an amount of money. They can lose part of this endowment on each trial with a probability of 1/3 if the true state is either right or left tilted (depending on the condition) – i.e., these monetary losses are not associated with participants' answers, but with the actual tilt of the sequence. In addition, participants are rewarded with a monetary bonus for correctly identifying sequences.

There are 3 treatment dimensions. Each participant will participate in each condition (a within-subject design) and the pattern recognition tasks are equally divided across treatments. The treatment dimensions are a) monetary loss is absent or present, and when it is present, it associated with patterns with a majority of red dots, or with a majority of blue dots, b) the incentives for accuracy are high or low, and c) the asymmetry in the number of blue and red dots differs. Our measure for ambiguity derives from a continuous variable of pattern difficulty that we dichotomize. The experiment uses a 3 x 2 x 4 within-subjects design with the factors loss pattern (no losses, associated with blue patterns, associated with red patterns), incentives (high, low) and ambiguity (high, medium high, medium low, low).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will run a linear regression of correct answers on the following variables
1. [loss pattern]: a dummy for whether the pattern is associated with a loss. A t-test on the coefficient tests our hypothesis on wishful thinking.
2. [high incentives1]: a dummy for the high incentive condition when there is no possibility of monetary losses. A t-test on the coefficient tests for the effect of high incentives in the absence of losses.
3. [high incentives2]: a dummy for the high incentive condition when there is a possibility of monetary losses. A t-test on the coefficient tests for the effect of high incentives when losses are present.
4. [interaction]: restricting ourselves to the pattern with losses, we will look at a dummy for the interaction of the loss patterns with the high incentive condition. A t-test on the coefficient for the interaction term tests whether the incentives reduce wishful thinking.
We cluster standard errors at the individual level. We run regressions where each observation is the individual average over trials in that condition, as well as regressions where we use all data points.

For patterns associated with a loss, incentives should raise performance both because of the overall effect and because of a reduction in wishful thinking. To test this joint effect, we conduct an additional t-test assessing whether high accuracy incentives improve performance on loss patterns only.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We conduct our online experiment on Prolific. We will exclude participants who fail to answer simple attention checks at the beginning and throughout the experiment.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit 400 subjects to complete the experiment.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will test whether self-reported concentration responds to incentives, as a manipulation check.
We will test whether self-reported anxiety responds to the presence of losses, as a manipulation check.

We will test our secondary hypothesis that wishful thinking increases with the ambiguity of the evidence.
Leveraging our within subject identification of wishful thinking, we will also study the heterogeneity of wishful thinking across subjects, correlating it with responses from the exit questionnaire.
We will test the robustness of our results if we exclude participants whose average accuracy is worse than chance.

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Wishful thinking in the gain and in the loss domain (#124703)

Created: 03/10/2023 12:33 PM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
In previous experiments we have found evidence for "wishful thinking". If a state of the world is associated with aversive outcomes, then subjects are less likely to correctly identify that state. Here, we will test whether wishful thinking is affected by the framing of the outcomes as (foregone) gains or losses.

3) Describe the key dependent variable(s) specifying how they will be measured.
Participants face 32 trials where they see a Gabor patch, and are asked to recognize whether it is tilted to the left or right. The main dependent variable is the accuracy on this task in each trial, i.e., whether the participant correctly identified the displayed patterns.

In particular, some patterns are associated with favorable and some with unfavorable outcomes. We are interested in "wishful thinking", defined as the difference in accuracy in recognizing patterns that result in favorable outcomes compared to the accuracy in recognizing patterns that result in unfavorable outcomes.

4) How many and which conditions will participants be assigned to?
We use a 2x2 mixed factorial design. The first treatment dimension varies which pattern (left- or right-tilted) results in a favorable outcome and which one results in an unfavorable outcome. This treatment varies within subjects: subjects face an equal number of trials where the left-titled pattern is favorable and where the right-tilted pattern is favorable.

The second treatment dimension varies the framing of the (un)favorable outcomes. In the loss treatment, participants are initially endowed with 16 pounds. On each trial, they incur a loss of 50 cents each time the unfavorable pattern occurs and incur no loss each time the favorable pattern occurs. In the gain treatment, participants are not endowed with money, but gain 50 cents each time the favorable pattern occurs and gain nothing each time the unfavorable pattern occurs. Note that these treatments result in an identical probability distribution over monetary outcomes. The frame treatment varies between subjects.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will run a linear regression where each trial is an observation. We regress a dummy for the correct answer (accuracy) on the following variables:

- Main effect of Wishful Thinking [unfavorable pattern]: a dummy for an unfavorable pattern (i.e. associated with a loss or absence of a gain). A t-test on the coefficient tests our hypothesis on wishful thinking in the gain frame (a negative coefficient indicates wishful thinking).

- Main effect of Loss Frame [Loss]: a dummy for the loss frame (relative to gain frame). A t-test on the coefficient tests our hypothesis that average accuracy differs across the two frames.

- Interaction between WT and Loss Frame [unfavorable pattern as loss]: a dummy for an unfavorable answer in the loss frame (i.e. associated with a loss). A t-test on the coefficient tests for the increase in wishful thinking in the loss vs. gain frame.

We cluster standard errors at the individual level. As robustness, we will run regressions where each observation is the individual average over trials in that condition. Regressions will control for the level of tilt, which differs between trials.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We conduct our online experiment on Prolific. We will exclude participants who fail to answer simple attention checks at the beginning and throughout the experiment.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will recruit 600 subjects to complete the experiment, 300 in each condition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will test whether wishful thinking is mediated by self-reported anxiety and by self-reported excitement, elicited with questionnaires after blocks of trials. We expect the former to play a larger role in the loss treatment and to be a more important driver of wishful thinking.
We vary the ambiguity of the pattern within-subject by using two different levels of the tilt and test whether wishful thinking increases with the ambiguity of the evidence.
We will test the robustness of our results if we exclude participants whose average accuracy is worse than chance or who indicated they found the instructions complex.

F Instructions, attention checks and subject exclusion

Instructions for the different Experiments can be found as part of the replication package on OSF: `https://doi.org/10.17605/OSF.IO/TZNPY`.

All experiments included quiz questions to check participant understanding of the instructions. In Experiments 2-5, the quiz questions are presented intermixed with the instructions, separated in two sections. Each question is repeated upon a wrong answer, and each section is repeated up to 2 times if 3 or more mistakes are made within that section. In Experiment 2 the first section was repeated up to 3 times, though this rarely happened. Participants were excluded from the experiment if they made more than 4 mistakes in total.

Several attention checks (see screen captures on the Online Supplementary Material: `https://doi.org/10.17605/OSF.IO/TZNPY`) were used at the beginning of the task for Experiment 2-5. Each was repeated upon a wrong response, and the experiment stopped at two total wrong responses. The last attention check (pressing a key written on screen) appeared once again at the end of the first block, and the experiment ended if answered incorrectly four times.