American Economic Review 2024, 114(4): 926–960
https://doi.org/10.1257/aer.20191068

Anticipatory Anxiety and Wishful Thinking†

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

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* C91, D12, D83, D91)

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, Shoulson, and Dorsey 2013; Ganguly and Tasoff 2016), while greater exposure to COVID-19 leads people to become more sanguine about the probability of infection (Orhun, Cohn, and Raymond 2021; Islam 2021). Populist politicians who promise easy fixes find more support in areas with weak economic prospects and declining growth rates (Mughan, Bean, and McAllister 2003; Obschonka et al. 2018).

*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.j.vanderweele@uva.nl). Pietro Ortoleva was the coeditor for this article. 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 an Ambizione grant (PZ00P3_174127). This study was approved by the Economics and Business Ethics Committee at the University of Amsterdam (refs. 20170510120541 and 20210202020244) and preregistered on aspredicted.org (15709, 57718, 83830, 89876, and 124703).
†Go to https://doi.org/10.1257/aer.20191068 to visit the article page for additional materials and author disclosure statements.

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 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 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 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). More 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

[1] For instance, consistent with wishful thinking, Oster, Shoulson, and Dorsey (2013) find that people at risk of Huntington's 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-19 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.

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 cannot 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, Taylor, and Martinez (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 roles 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, Norton, and Weber 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, Taylor, and Martinez (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]

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

[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 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, Korn, and Dolan 2011; Sharot et al. 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).

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 US$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, Tripodi, and van der Weele (2022) find no evidence for an effect on self-persuasion and polarization in a debating context. Here, we 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 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, Fernbach, and Hagmayer 2010; Chance and Norton 2015; Gino, Norton, and Weber 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 Appendixes E and F.

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

[4] Buser, Gerhards, and van der Weele (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.

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 desensitization (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 predetermined 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 average 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.

[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).

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	£1	£0 or £1
Task difficulty levels	Tilt size (3 levels)	Tilt size (2 levels)	Likelihood ratio (continuous)	Dot ratio (4 levels)
Accuracy bonus levels	€1 €20	£0.10 £8	£0.05 £10	£0.05 £10
Other design elements	Confidence measure Replication exp.		Treatment reminders	Treatment reminders Self-timed task
Start/end date	November 12, 2018 December 5, 2018	February 23, 2021 February 29, 2021	January 3, 2022 January 4, 2022	March 8, 2022 March 8, 2022
Location	CREED Laboratory (Amsterdam)	Online (Prolific)	Online (Prolific)	Online (Prolific)

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 experiment in turn.

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]

The visual task was to determine whether a grating (Gabor patch) was tilted toward 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

[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, Meyer, Fehr, and Ruff 2015; Engelmann, Meyer, Ruff, and Fehr 2019).

[7] The wrist of the participant's nondominant hand was connected to a Digitimer DS5 isolated bipolar current stimulator, which itself was connected to MATLAB through National Instruments USB x-series. A participant induced themselves with a series of shocks, which they 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, Meyer, Fehr, and Ruff 2015; Engelmann, Meyer, Ruff, and Fehr 2019).

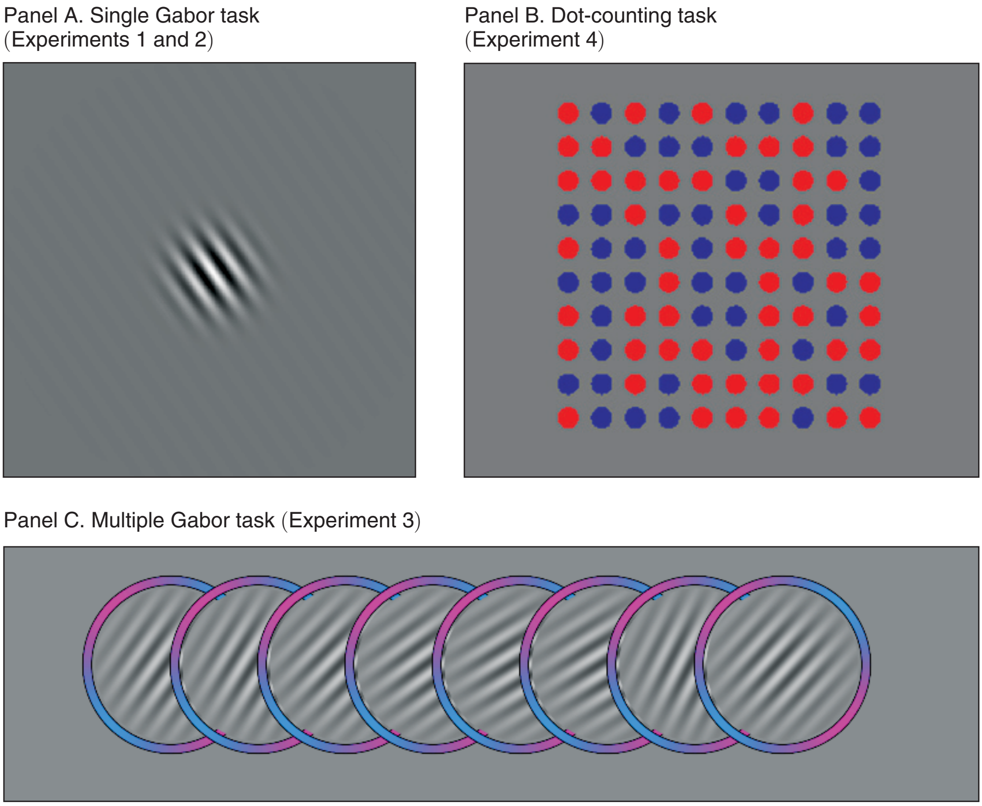

Figure 1. Examples of the Visual Tasks in the Various Experiments

(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 percent (completely uncertain) to 100 percent (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 participant's risk preferences.[8]

Next, participants faced an anticipation screen (2,000–8,000ms), asking them to wait for the shock resolution. Finally, the electric shock was administered or not (1,000ms). 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 3 sessions, each divided into 4 blocks of 18 trials. The four blocks correspond to four conditions of a 2×2 factorial design (Shock ×

[8] Subjects indicate their subjective probability $x \in \{50, 55, \ldots, 95, 100\}$ 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, Tremewan, and van der Weele (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.

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. 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 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 for a session that lasted on average slightly over an hour.

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 F). 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 percent 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 3 parts of up to 32 trials. If the participant ran out of endowment before the 32 trials, then the remaining trials were canceled.[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 endowment with £5 losses (*High Loss* condition), a

[9] The three difficulty levels were calibrated to result in accuracy levels of 60 percent, 70 percent, and 80 percent. 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 recalibrated 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.

[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 (i) the loss will likely not count and (ii) 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.

£0.50 endowment with £0.10 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 VIIC 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 or £0.10. Unlike in the previous experiment, we did not elicit confidence measures. Instead, we randomly selected one £8 trial and one £0.10 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 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 versus right) and the incentive for accuracy were varied at the block level, where a block consisted of eight 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. *Experiments 3 and 4: Task Characteristics and Incentive Effects*

Besides 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 a 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]

[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.”

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 eight tilted Gabor patches spaced over four seconds, as illustrated in Figure 1. The tilts were generated from one of two distributions of patterns that were biased toward either left- or right-leaning patterns. We then asked participants to infer which distribution generated the patterns, and define a correct 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 from which they would lose £1 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 (£0.05 versus £10) and the orientation of the loss patterns (left versus 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, Gietl, and Heinemann (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 from which they would lose €1 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 (£0.05 versus £10) and the color of the loss pattern (blue versus 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 or losses, the order of which was randomized to be either before or after the

[13] Occasionally, 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.

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 off 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 $dp(r_\theta = 1, s)/ds > 0$ and $dp(r_\theta = 0, s)/ds < 0$. 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-Marschak 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

$$\sigma_z\left[r_z\hat{p} + (1 - r_z)(1 - \hat{p})\right]qZ.$$

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

[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.

not only experience the disutility of anticipatory anxiety but also the disutility of actually receiving the shock, which is given by $\left[r_z p + (1 - r_z)(1 - p)\right]qZ$.

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

$$U = \frac{1}{2}(1 + 2p\hat{p} - \hat{p}^2)M - \left[r_z p + (1 - r_z)(1 - p)\right]qZ$$
$$- \sigma_z\left[r_z\hat{p} + (1 - r_z)(1 - \hat{p})\right]qZ - \lambda(s)(p - \hat{p})^2.$$

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.,* $dW/ds < 0$.

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.*, $dW/dM < 0$.

Experiment 2 varies psychological stakes by varying the loss associated with a loss pattern. By relabeling 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.*, $dW/dZ > 0$.

Online Appendix D features a number of extensions of the model. In online Appendix 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 online Appendix 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 online Appendix 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 confirm some predictions of the model and are at odds with some others. To capture these discrepancies, online Appendix 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 4 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

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)
Dependent variable	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 × HAB		0.944		0.994		−0.110		1.363
		(1.787)		(1.771)		(0.881)		(1.325)
Shock/Loss pattern × 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 × 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	3,415	3,415	3,408	3,408	6,502	6,502
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 are clustered by individual.

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 IVC 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

[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.

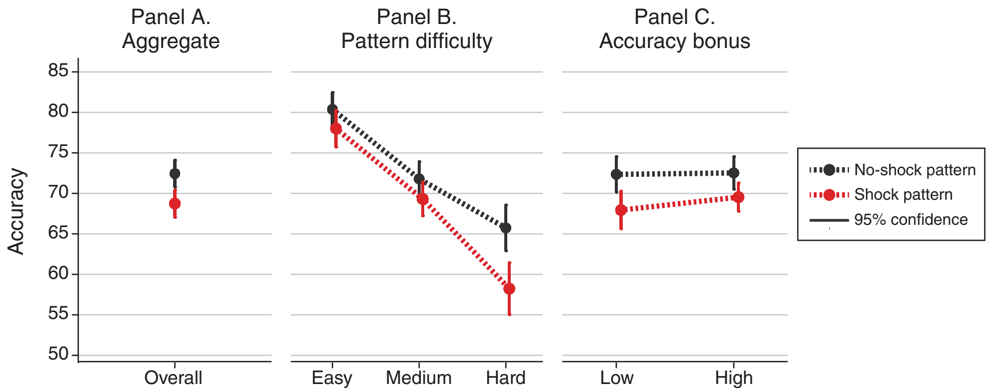

Figure 2. Electric Shocks and Accuracy in Experiment 1

Notes: Average accuracy levels, split by shock and no-shock pattern. Bars indicate 95 percent 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.

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]

Nevertheless, as an external benchmark, one might consider mammogram reading, a complex pattern recognition task with a high-stakes 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 4

[16] It is hard to pinpoint the differences in effect sizes. Relative to Experiment 1, Experiments 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 who were generally more comfortable with the relevant source of anticipatory anxiety. Experiments 3 and 4 further differ in the perceptual task and other implementation details.

[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.

percentage points (72.3 versus 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, online Appendix 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 of Figure 2 displays the impact of raising the prize for the BDM mechanism from €1 to €20. 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 percent confidence in the wrong answer) to 100 (meaning the subject indicated 100 percent 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 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 eight trials in which losses could occur. Figure 3 shows the density of different anxiety ratings in the Low Loss (£0.10) and High Loss (£5) 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 given category, so $N = 221$ in each category. Table 2, columns 3 and 4 provide regression evidence associated with these results, and online Appendix Table A.6 provides robustness across regression models. Results

[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).

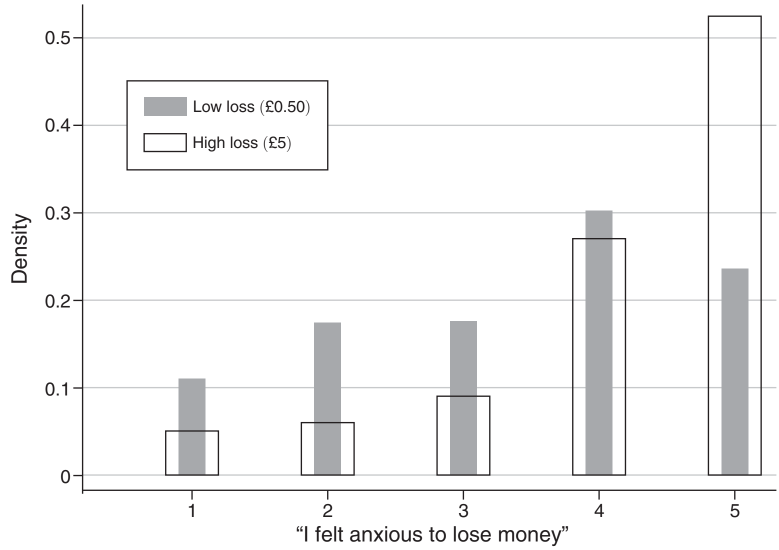

FIGURE 3. MANIPULATION CHECK

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

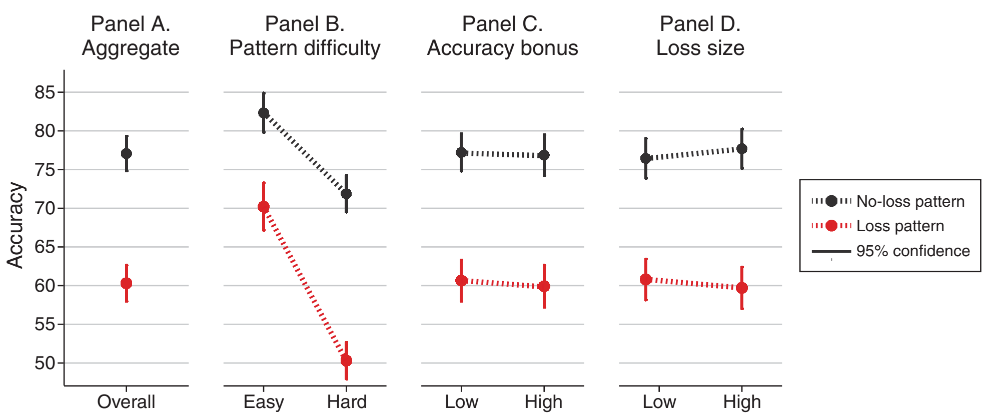

FIGURE 4. MONETARY LOSSES AND ACCURACY IN EXPERIMENT 2

Notes: Average accuracy levels, split by loss and no-loss patterns. Bars indicate 95 percent 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.

exclude the Neutral condition since this is not a test of wishful thinking and is discussed in Section VIIC.

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, 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 £0.10 to £5. 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.

Panel (i) of Figure 5 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, respectively. Again, we find little evidence that incentives improve performance: A high bonus improves accuracy by about 0.7 percentage point, but the effect is not close to being statistically significant. Moreover, there is no interaction with the loss pattern, so no reduction in wishful thinking from higher accuracy incentives.

Panel (ii) of Figure 5 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 as an interaction with wishful thinking. Panel C shows the pattern for the different levels of the accuracy bonus of £0.05 and £10. Unlike for the tasks we considered above, incentives improve performance: moving from the low to the high bonus improves accuracy by about 1.6 percentage point ($p < 0.01$; 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.

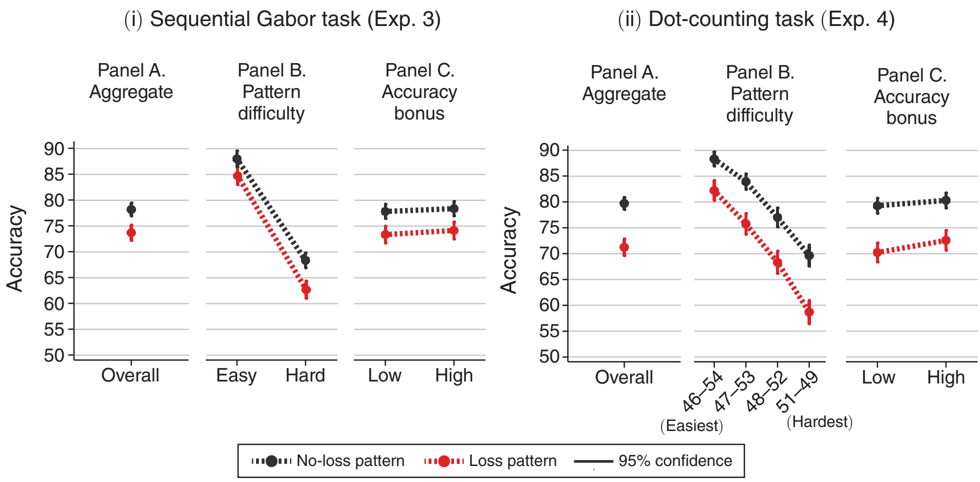

FIGURE 5. ACCURACY IN THE MULTIPLE GABOR AND DOT-COUNTING TASKS IN EXPERIMENTS 3 AND 4

Notes: Average accuracy levels, split by loss and no-loss pattern. Bars indicate 95 percent 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.

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] Overall, the data do not present a coherent story but suggest that repeated shocks may lead to desensitization, while monetary losses can lead to heightened wishful thinking in some contexts. In particular, wishful thinking in the first half of Experiment 1 is more than twice as large as in the second half. Although statistically nonsignificant ($p = 0.102$), 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

[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 analyzes 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 the participant has gone through. In a second set of analyses, we simply compare wishful thinking in the first half and the second half of the experiment.

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, Gerhards, and van der Weele 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 online Appendix Table B.1.

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

[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.

[22] Our measure is based on the defensive pessimism questionnaire (Norem 2008). Following Lim (2009), we focus on the pessimism subscale, 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.

Table 3—Emotional and Cognitive Covariates of Wishful Thinking

Dependent variable:	Wishful thinking (1)	Wishful thinking (2)	Wishful thinking (3)	Wishful thinking (4)
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	1,050	625	744	422
R^2	0.066	0.053	0.086	0.076

Notes: OLS regressions of wishful thinking on emotional and 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 seven-point Likert scale from very easy to very difficult. All regressions contain experiment dummies. Standard errors in parentheses.

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 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

[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

correlations and exclude participants who reported difficulties following the 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 motivated cognition. 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 £0.87.[25] This corresponds to roughly ten 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.

tailor-made measure. However, perhaps unsurprisingly, the positive correlation between BAI and wishful thinking is not statistically significant.

[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 toward 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 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 percent. So across trials, associating the true state of the world with an anxiety-inducing outcome led to a 10.9 percentage point decrease in accuracy and an expected loss of $0.109 \times 8 = 0.87$ pounds. For the most ambiguous patterns, the decrease in accuracy is 12.8 percentage points, and the expected loss is £1.02.

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 (0.821)	0.905 (0.905)	1.610 (0.574)	1.152 (0.690)	0.0433 (0.0125)	0.159 (0.0179)	0.145 (0.0340)	0.204 (0.0229)
Experiment no.	2	2	4	4	2	4	2	4
Conditions	Losses-present	Neutral	Losses-present	Neutral	All	All	All	All
Observations	11,396	7,072	21,114	13,024	18,468	34,133	18,468	34,138
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 a regression across all conditions/experiments where the outcome variable is log response time in each trial, measured in milliseconds. Column 6 shows a regression across all Experiments 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 a five-point Likert scale. Regressions control for pattern difficulty and include a constant (not reported). Standard errors in parentheses are clustered by individual.

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 conditions in Experiments 2 and 4 that do not feature the threat of a loss. Table 4 shows OLS regressions of the impact of the accuracy bonus on accuracy in Experiments 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 consider their answer.[26] Columns 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 eight-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

[26] Response times are routinely used in cognitive science and economics as a measure of cognitive effort (e.g., Bettman, Johnson, and Payne 1990; Camerer and Hogarth 1999; Enke et al. 2021) and are an important component of recent theories of decision-making (Fudenberg, Strack, and Strzalecki 2018; Alós-Ferrer, Fehr, and Netzer 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).

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.

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 shows a statistically significant increase 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 postexperimental questionnaire whether they counted dots. We find that 9 percent of subjects replied "Always," 38 percent replied "Sometimes," and 53 percent 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 noncounters 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 shows 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 percent of the time (versus 73 percent for the

[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.

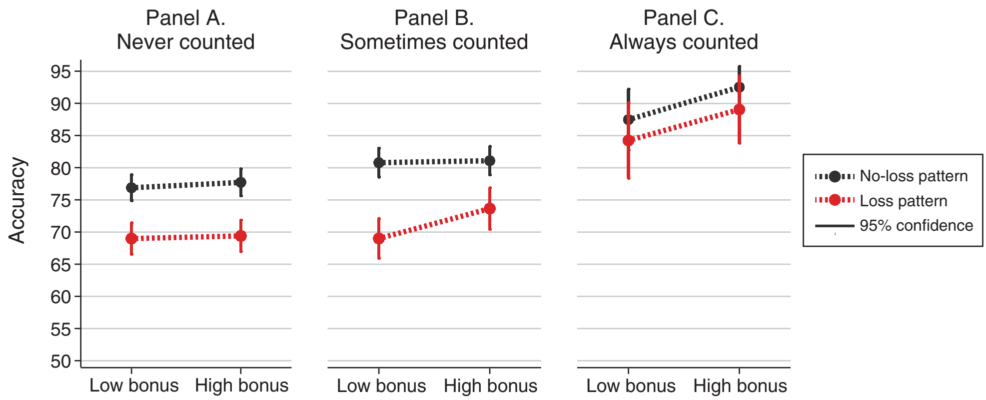

Figure 6. Accuracy in the Dot-Counting Task

Notes: Average accuracy levels, split by loss and no-loss patterns. Bars indicate 95 percent 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 ($N = 36$).

"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 IVE, 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 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 previous work has shown that shocks activate feelings of anxiety and fear (see Section IIB). 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 IVE. 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 £0.50 if the loss pattern appeared on the screen from an initial endowment of £16. In the Gain Frame treatment, subjects gained £0.50 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. 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 £1. After each block of eight trials, we asked subjects about their experienced anxiety and their excitement about whether or not they would lose/gain money, measured on a five-point Likert scale. We recruited 300 participants for each (between-subject) treatment on Prolific. 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).

To further investigate the reverse effect in the domain of gains, we look at two moderators: (i) high levels of self-reported emotions and (ii) 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

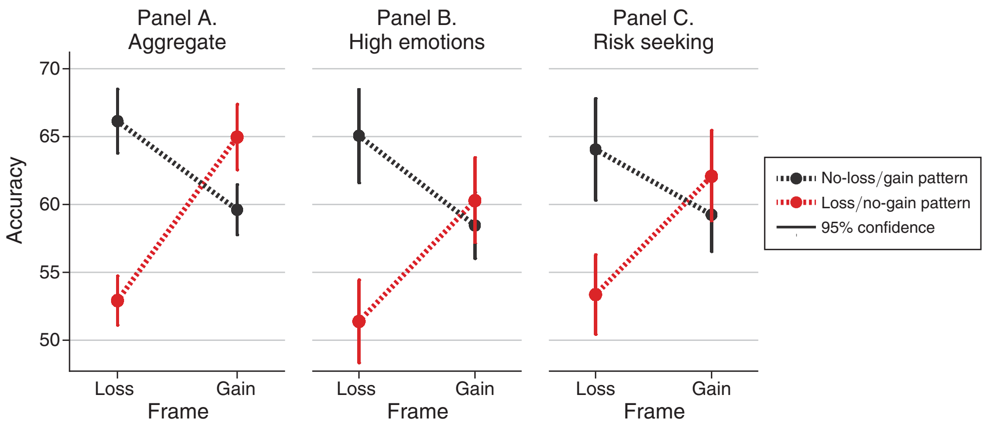

Figure 7. Accuracy in the Gain Frame and Loss Frame

Notes: Average accuracy levels, split by loss (no-gain) and no-loss (gain) pattern. Bars indicate 95 percent 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)?"

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. Online Appendix 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 that 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 Experiments 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 Tables 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 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. 1982. "The Economic Consequences of Cognitive Dissonance." *American Economic Review* 72 (3): 307–19.

Alós-Ferrer, Carlos, Ernst Fehr, and Nick Netzer. 2021. "Time will Tell: Recovering Preferences when Choices are Noisy." *Journal of Political Economy* 129 (6): 1828–77.

Armor, David A., and Aaron M. Sackett. 2006. "Accuracy, Error, and Bias in Predictions for Real Versus Hypothetical Events." *Journal of Personality and Social Psychology* 91 (4): 583–600.

Auriol, Emmanuelle, Julie Lassebie, Amma Panin, Eva Raiber, and Paul Seabright. 2020. "God Insures Those Who Pay? Formal Insurance and Religious Offerings in Ghana." *Quarterly Journal of Economics* 135 (4): 1799–1848.

Balcetis, Emily, and David Dunning. 2006. "See What You Want to See: Motivational Influences on Visual Perception." *Journal of Personality and Social Psychology* 91 (4): 612–25.

[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; Kőszegi 2010), occupational choice and the labor market equilibrium (Akerlof and Dickens 1982; Santos-Pinto, Dell, and Opromolla 2018), information acquisition (Yariv 2002; Eliaz and Spiegler 2006; Loewenstein 2006), principal-agent communication (Kőszegi 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).

Barron, Kai. 2021. "Belief Updating: Does the 'Good-News, Bad-News' Asymmetry Extend to Purely Financial Domains?" *Experimental Economics* 24 (1): 31–58.
Bénabou, Roland. 2008. "Ideology." *Journal of the European Economic Association* 6 (2-3): 321–52.
Bénabou, Roland. 2013. "Groupthink: Collective Delusions in Organizations and Markets." *Review of Economic Studies* 80 (2): 429–62.
Bénabou, Roland, and Jean Tirole. 2002. "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics* 117 (3): 871–915.
Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–55.
Bentham, Jeremy. 1789. *An Introduction to the Principles of Morals and Legislation.* Oxford: Clarendon Press.
Berns, Gregory S., Jonathan Chappelow, Milos Cekic, Caroline F. Zink, Giuseppe Pagnoni, and Megan E. Martin-Skurski. 2006. "Neurobiological Substrates of Dread." *Science* 754 (5774): 754–58.
Bettman, James R., Eric J. Johnson, and John W. Payne. 1990. "A Componential Analysis of Cognitive Effort in Choice." *Organizational Behavior and Human Decision Processes* 45 (1): 111–39.
Bosch-Rosa, Ciril, Daniel Gietl, and Frank Heinemann. 2021. "Risk-Taking under Limited Liability: Quantifying the Role of Motivated Beliefs." Unpublished.
Bracha, Anat, and Donald J. Brown. 2012. "Affective Decision Making: A Theory of Optimism Bias." *Games And Economic Behavior* 75 (1): 67–80.
Bridet, Luc, and Peter Schwardmann. 2023. "Selling Dreams: Endogenous Optimism in Lending Markets." Unpublished.
Brunnermeier, Markus K., and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118.
Burton, Jason W., Adam J. L. Harris, Punit Shah, and Ulrike Hahn. 2022. "Optimism Where There is None: Asymmetric Belief Updating Observed with Valence-Neutral Life Events." *Cognition* 218: 104939.
Buser, Thomas, Leonie Gerhards, and Joël J. van der Weele. 2018. "Responsiveness to Feedback as a Personal Trait." *Journal of Risk and Uncertainty* 56 (2): 165–92.
Camerer, Colin F., and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19 (1): 7–42.
Caplin, Andrew, and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics* 116 (1): 55–79.
Caplin, Andrew, and John Leahy. 2004. "The Supply of Information by a Concerned Expert." *Economic Journal* 114 (497): 487–505.
Caplin, Andrew, and Mark Dean. 2014. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." NBER Working Paper 19876.
Chance, Zoe, and Michael I. Norton. 2015. "The What and Why of Self-Deception." *Current Opinion in Psychology* 6: 104–07.
Clithero, John A. 2018. "Response Times in Economics: Looking Through the Lens of Sequential Sampling Models." *Journal of Economic Psychology* 69: 61–86.
Coutts, Alexander. 2019. "Testing Models of Belief Bias: An Experiment." *Games and Economic Behavior* 113: 549–65.
Dean, Mark, and Nate Leigh Neligh. 2019. "Experimental Tests of Rational Inattention." Unpublished.
Dewan, Ambuj, and Nathaniel Neligh. 2020. "Estimating Information Cost Functions in Models of Rational Inattention." *Journal of Economic Theory* 187: 105011.
Drobner, Christoph. 2022. "Motivated Beliefs and Anticipation of Uncertainty Resolution." *American Economic Review: Insights* 4 (1): 89–105.
Drugowitsch, Jan, Valentin Wyart, Anne-Dominique Devauchelle, and Etienne Koechlin. 2016. "Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality." *Neuron* 92 (6): 1398–1411.
Dunning, David, and Emily Balcetis. 2013. "Wishful Seeing: How Preferences Shape Visual Perception." *Journal of Experimental Social Psychology* 22 (1): 33–37.
Eil, David, and Justin M. Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information About Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38.
Eliaz, Kfir, and Ran Spiegler. 2006. "Can Anticipatory Feelings Explain Anomalous Choices of Information Sources?" *Games and Economic Behavior* 56 (1): 87–104.
Engelmann, Jan B., Maël Lebreton, Nahuel A. Salem-Garcia, Peter Schwardmann, and Joël J. van der Weele. 2024. "Replication data for: Anticipatory Anxiety and Wishful Thinking." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E195781V1.

Engelmann, Jan B., Friederike Meyer, Christian C. Ruff, and Ernst Fehr. 2019. "The Neural Circuitry of Affect-Induced Distortions of Trust." *Science Advances* 5 (3): 3413.

Engelmann, Jan B., Friederike Meyer, Ernst Fehr, and Christian C. Ruff. 2015. "Anticipatory Anxiety Disrupts Neural Valuation during Risky Choice." *Journal of Neuroscience* 35 (7): 3085–99.

Enke, Benjamin, Uri Gneezy, Brian Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven. 2021. "Cognitive Biases: Mistakes or Missing Stakes?" *Review of Economics and Statistics* 105 (4): 1–45.

Eyal, Peer, Rothschild David, Gordon Andrew, Evernden Zak, and Damer Ekaterina. 2021. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods* 54: 1642–62.

Findling, Charles, and Valentin Wyart. 2021. "Computation Noise in Human Learning and Decision-Making: Origin, Impact, Function." *Current Opinion in Behavioral Sciences* 38: 124–32.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review* 108 (12): 3651–84.

Ganguly, Ananda, and Joshua Tasoff. 2016. "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future." *Management Science* 63 (12): 4037–60.

Garbers, Yvonne, and Udo Konradt. 2014. "The Effect of Financial Incentives on Performance: A Quantitative Review of Individual and Team-Based Financial Incentives." *Journal of Occupational and Organizational Psychology* 87 (1): 102–37.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212.

Grillon, Christian. 2008. "Models and Mechanisms of Anxiety: Evidence from Startle Studies." *Psychopharmacology* 199 (3): 421–37.

Grossman, Zachary, and Joël J. van der Weele. 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15 (1): 173–217.

Hadjiiski, Lubomir, Heang-Ping Chan, Berkman Sahiner, Mark A. Helvie, Marilyn A. Roubidoux, Caroline Blane, Chintana Paramagul, et al. 2004. "Improvement in radiologists' Characterization of Malignant and Benign Breast Masses on Serial Mammograms with Computer-Aided Diagnosis: An ROC Study." *Radiology* 233 (1): 255–65.

Haisley, Emily C., and Roberto A. Weber. 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68 (2): 614–25.

Houssami, Nehmat, Les Irwig, Judy M. Simpson, Merran McKessar, Steven Blome, and Jennie Noakes. 2004. "The Influence of Clinical Information on the Accuracy of Diagnostic Mammography." *Breast Cancer Research and Treatment* 85 (3): 223–28.

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

Islam, Marco. 2021. "Motivated Risk Assessments." Unpublished.

Jevons, William S. 1879. *The Theory of Political Economy.* London: Macmillan and Company.

Kappes, Andreas, and Tali Sharot. 2019. "The Automatic Nature of Motivated Belief Updating." *Behavioural Public Policy* 3 (1): 87–103.

Kőszegi, Botond. 2006. "Emotional Agency." *Quarterly Journal of Economics* 121 (1): 121–55.

Kőszegi, Botond. 2010. "Utility from Anticipation and Personal Equilibrium." *Economic Theory* 44 (3): 415–44.

Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-Dependent Consumption Plans." *American Economic Review* 99 (3): 909–36.

Krizan, Zlatan, and Paul D. Windschitl. 2007. "The Influence of Outcome Desirability on Optimism." *Psychological Bulletin* 133 (1): 95–121.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3): 480–98.

Lench, Heather C., and Peter H. Ditto. 2008. "Automatic Optimism: Biased Use of Base Rate Information for Positive and Negative Events." *Journal of Experimental Social Psychology* 44 (3): 631–39.

Leong, Yuan Chang, Brent L. Hughes, Yiyu Wang, and Jamil Zaki. 2019. "Neurocomputational Mechanisms Underlying Motivated Seeing." *Nature Human Behaviour* 3 (9): 962–73.

Lerman, Caryn, Chanita Hughes, Stephen J. Lemon, David Main, Carrie Snyder, Carolyn Durham, Steven Narod, and Henry T. Lynch. 1998. "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* 16 (5): 1650–54.

Levy, Raphaël. 2014. "Soothing Politics." *Journal of Public Economics* 120: 126–33.

Le Yaouanq, Yves. 2023. "A Model of Voting with Motivated Beliefs." *Journal of Economic Behavior and Organization* 213: 394–408.

Lim, Lena. 2009. "A Two-Factor Model of Defensive Pessimism and its Relations with Achievement Motives." *Journal of Psychology* 143 (3): 318–36.

Loewenstein, George. 1987. "Anticipation and the Valuation of Delayed Consumption." *Economic Journal* 97 (387): 666–84.
Loewenstein, George. 2006. "The Pleasures and Pains of Information." *Science* 312 (5774): 704–06.
Loewenstein, George, and Andras Molnar. 2018. "The Renaissance of Belief-Based Utility in Economics." *Nature Human Behaviour* 2 (3): 166–67.
Macera, Rosario. 2014. "Dynamic Beliefs." *Games and Economic Behavior* 87: 1–18.
Mayraz, Guy. 2011. "Wishful Thinking." Unpublished.
Melnikoff, David E., and Nina Strohminger. 2020. "The Automatic Influence of Advocacy on Lawyers and Novices." *Nature Human Behaviour* 4 (12): 1258–64.
Mijović-Prelec, Danica, and Drazen Prelec. 2010. "Self-Deception as Self-Signalling: A Model and Experimental Evidence." *Phil. Trans. of the Royal Society B* 365 (1538): 227–40.
Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2022. "Managing Self-Confidence: Theory and Experimental Evidence." *Management Science* 68 (11): 7793–7817.
Molnar, Andras, and George Loewenstein. 2021. "Thoughts and Players: An Introduction to Old and New Economic Perspectives on Beliefs." In *The Cognitive Science of Beliefs: A Multidisciplinary Approach,* edited by Julien Musolino, Joseph Sommer, and Pernille Hemmer, 321–50. Cambridge: Cambridge University Press.
Mughan, A., C. Bean, and I. McAllister. 2003. "Economic Globalization, Job Insecurity and the Populist Reaction." *Electoral Studies* 22 (4): 617–33.
Norem, Julie. 2008. *The Positive Power of Negative Thinking.* New York, NY: Basic Books.
Norem, Julie K., and Nancy Cantor. 1986. "Defensive Pessimism: Harnessing Anxiety as Motivation." *Journal of Personality and Social Psychology* 51 (6): 1208–17.
Obschonka, Martin, Michael Stuetzer, Peter J. Rentfrow, Neil Lee, Jeff Potter, and Samuel D. Gosling. 2018. "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* 9 (3): 285–98.
Orhun, A. Yesim, Alain Cohn, and Collin Raymond. 2021. "Motivated Optimism and Workplace Risk." Unpublished.
Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review* 103 (2): 804–30.
Pronk, Thomas, Dylan Molenaar, Reinout W. Wiers, and Jaap Murre. 2021. "Methods to Split Cognitive Task Data for Estimating Split-Half Reliability: A Comprehensive Review and Systematic Assessment." *Psychonomic Bulletin & Review* 29 (1): 44–54.
Saccardo, Silvia, and Marta Serra-Garcia. 2023. "Enabling or Limiting Cognitive Flexibility? Evidence of Demand for Moral Commitment." *American Economic Review* 113 (2): 396–429.
Salvador, Alexandre, Luc H. Arnal, Fabien Vinckier, Philippe Domenech, Raphaël Gaillard, and Valentin Wyart. 2022. "Premature Commitment to Uncertain Decisions during Human NMDA Receptor Hypofunction." *Nature Communications* 13 (1): 1–15.
Santos-Pinto, Luís, Michele Dell, and Luca David Opromolla. 2018. "A General Equilibrium Theory of Occupational Choice under Optimistic Expectations." Unpublished.
Schlag, Karl H., James Tremewan, and Joël J. van der Weele. 2015. "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs." *Experimental Economics* 18 (3): 457–90.
Schmitz, Anja, and Christian Grillon. 2012. "Assessing Fear and Anxiety in Humans using the Threat of Predictable and Unpredictable Aversive Events (the NPU-threat test)." *Nature Protocols* 7 (3): 527–32.
Schwardmann, Peter. 2019. "Motivated Health Risk Denial and Preventative Health Care Investments." *Journal of Health Economics* 65: 78–92.
Schwardmann, Peter, and Joël van der Weele. 2019. "Deception and Self-Deception." *Nature Human Behaviour* 3 (10): 1055–61.
Schwardmann, Peter, Egon Tripodi, and Joël J. van der Weele. 2022. "Self-Persuasion: Evidence from Field Experiments at International Debating Competitions." *American Economic Review* 112 (4): 1118–46.
Shah, Punit, Adam J. L. Harris, Geoffrey Bird, Caroline Catmur, and Ulrike Hahn. 2016. "A Pessimistic View of Optimistic Belief Updating." *Cognitive Psychology* 90: 71–127.
Sharot, Tali, Christoph W. Korn, and Raymond J. Dolan. 2011. "How Unrealistic Optimism is Maintained in the Face of Reality." *Nature Neuroscience* 14 (11): 1475–79.
Sharot, Tali, Marc Guitart-Masip, Christoph W. Korn, Rumana Chowdhury, and Raymond J. Dolan. 2012. "How Dopamine Enhances an Optimism Bias in Humans." *Current Biology* 22 (16): 1477–81.
Simmons, Joseph P., and Cade Massey. 2012. "Is Optimism Real?" *Journal of Experimental Psychology: General* 141 (4): 630–34.

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

Sinding Bentzen, Jeanet. 2021. "In Crisis, We Pray: Religiosity and the COVID-19 Pandemic." *Journal of Economic Behavior & Organization* 192: 541–83.

Sloman, Steven A., Philip M. Fernbach, and York Hagmayer. 2010. "Self-Deception Requires Vagueness." *Cognition* 115 (2): 268–81.

Wyart, Valentin, and Etienne Koechlin. 2016. "Choice Variability and Suboptimality in Uncertain Environments." *Current Opinion in Behavioral Sciences* 11: 109–15.

Yariv, Leeat. 2002. "I'll See It When I Believe It - A Simple Model of Cognitive Consistency." Unpublished.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–63.