Anticipatory Anxiety and Wishful Thinking*

Jan B. Engelmann[1], Maël Lebreton[2], Nahuel A. Salem-Garcia[3], Peter Schwardmann[4], and Joël J. van der Weele[1]

[1]University of Amsterdam, Tinbergen Institute

[2]Paris School of Economics

[3]University of Geneva

[4]Carnegie Mellon University

July 25, 2022

Abstract

We test the hypothesis that anxiety about adverse future outcomes leads to wishful thinking. Across four experiments (N=1,116), participants perform pattern recognition tasks in which some patterns may result in an electric shock or a monetary loss. Participants engage in significant wishful thinking, as they are less likely to correctly identify patterns that may lead to a shock or loss. Wishful thinking increases with greater ambiguity of the visual evidence and is only disciplined by higher accuracy incentives when accuracy depends on participants' cognitive effort. Wishful thinking is heterogeneous across and stable within individuals.

JEL classification: C91, D83

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

*The authors thank the editor, four anonymous referees, Douglas Bernheim, Andrew Caplin, Mark Dean, Yves Le Yaouanq, George Loewenstein, Nathaniel Neligh, Matthew Rabin, Simeon Schudy, Claudia Senik, Severine Toussaert, and seminar participants at the 2021 AEA Meeting, Stanford University, Amsterdam Brain and Cognition Center, ECBE, University of Chicago, New York University, MBEES workshop, MiddExLab Seminar, Belief Based Utility Conferences at BRIQ and UvA, Paris School of Economics, LMU Munich, and UC Santa Barbara for useful comments. Li-Ang Chang, Sára Khayouti and Nik Mautner Markhof provided excellent research assistance. For financial support, Peter Schwardmann thanks the Bavarian Academy of Sciences and Humanities, Joël van der Weele thanks the Dutch Science Association (NWO) via a personal VIDI grant (452-17-004), and Maël Lebreton thanks the Swiss National Science Foundation (SNSF) via a personal Ambizione grant (PZ00P3$_1$74127).

1 Introduction

Many commonly held beliefs seem to be inspired by their comforting properties rather than their realism. Billions of adherents of the major religions believe in an afterlife, despite a lack of evidence 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., 2017). People at risk of serious diseases avoid medical testing and remain optimistic about their health status (Lerman et al., 1998; Oster et al., 2013; Ganguly and Tasoff, 2016), while greater exposure to Covid-19 leads people to become more sanguine about the probability of infection (Orhun et al., 2021; Islam, 2021). Populist politicians that promise easy fixes find more support in areas with weak economic prospects and declining growth rates (Mughan et al., 2003; Obschonka et al., 2018).

These findings suggest that the adoption of comforting beliefs affects important decisions and originates in anxiety about adverse future outcomes. At the same time, it is hard to establish with field data that beliefs are the product of self-deception or wishful thinking, or to determine the precise motives behind self-deception.[1] Laboratory experiments on wishful thinking or optimism bias have also not established a clear causal link with anxiety, due to a number of factors that we discuss in the next section. This leaves an important gap in the literature: pinning down the motives and processes behind wishful thinking allows predicting its occurrence as well as the design of strategies that may abate it, such as tackling the underlying anxiety, providing precise information, or raising the material costs of false beliefs.

To make progress on these matters we study the effect of experimentally induced anxiety on belief formation in a tightly controlled setting. Our four preregistered experiments (combined $N = 1,116$) incentivize participants to correctly identify which of two types of patterns they see on their screen. We 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, a proven method of inducing anxiety. In our second, third, and fourth experiment, the adverse outcome is a monetary loss. Since participants have no control over the occurrence of these adverse outcomes, the payoff-maximizing strategy is to identify the patterns as accurately as possible. By contrast, anticipatory anxiety about the shock or loss may cause wishful thinking, a

[1] For instance, consistent with wishful thinking, Oster et al. (2013) find that people at risk of Huntington disease are optimistic before they get tested for the disease, but are reluctant to test, especially when they have low objective risk. However, without exogenous variation in the motives to hold optimistic beliefs, it neither clear whether initial optimism is the result of wishful thinking nor whether it is driven by a desire to avoid feeling anxious. Furthermore, Islam (2021) finds that individuals self-deceive about the risk of a Covid infection in deciding whether to go to a coffee shop during the pandemic. At the same time, they distort beliefs about the risk for others rather than for themselves, suggesting that self-deception is driven by social motives rather than anxiety about one's own health.

belief that the anxiety-inducing state of the world is less likely than it really is. In our experiment, wishful thinkers will then 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 distorts the visual signal to optimally trade off the anticipatory utility benefits from alleviated anxiety and the material costs stemming from inferior decision making. Apart from wishful thinking or overoptimism, this generates a number of additional predictions, most notably about the role of ambiguity and accuracy incentives, which are shared by most models motivated cognition. We experimentally test these predictions across our experiments, and develop an individual measure of wishful thinking of our participants.

The results show clear evidence for wishful thinking. In all experiments, participants are significantly less accurate in identifying patterns that may lead to an adverse outcome. This result obtains for different sources of anxiety (shock versus monetary loss), different pattern identification tasks, and in different settings (the experiment taking place online versus in the laboratory). Our dataset can rule out competing explanations for the observed effect, like illusions 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. We also find that, on balance, wishful thinking does not depreciate and remains high in later trials of the experiments. These results are strikingly robust in comparison to the literature in economics and psychology that we review below, which has yielded mixed findings, mostly in the domain of positive outcomes.

In all experiments, we find that wishful thinking is more pronounced when the evidence is more ambiguous - i.e. when the different patterns are difficult to distinguish. This is true across three different visual inference tasks in which we vary ambiguity in distinct ways. Therefore, in line with previous results (Haisley and Weber, 2010; Sloman et al., 2010) and recent theoretical work (Caplin and Leahy, 2019), we find that the precision of the signal constrains motivated belief formation.

We also test whether wishful thinking decreases with increasing material costs of false beliefs. We vary the accuracy bonus that participants can earn from a correct answer by factors 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. This is not for lack of trying on behalf of the participants, as response times and self-reported concentration increase significantly under higher incentives. In Experiment 2, we also vary the magnitude of monetary losses, i.e. the adverse outcome that participants may feel anxious about or the "anticipatory payoffs". Similar to our results on material incentives, we find that increasing losses increases self-reported anxiety but has no significant effect on wishful thinking.

We interpret these null results in the context of our model, where the agent first observes a signal and then distorts her mental representation of it to become more optimistic. The results of our first three experiments imply that participants engage in this ex-post signal distortion, but that the extent of the distortion does not respond to either material or anticipatory payoffs at the margin. However, our results suggest another way in which wishful thinking may respond to accuracy incentives. Incentives increase participants' ex-ante effort to form an accurate representation. If these efforts are successful, they may make wishful thinking harder in the same way that being shown a less ambiguous pattern increases accuracy and reduces the scope for wishful thinking in our experiments. In our first three experiments, this mechanism is ruled out by the fact that accuracy is insensitive to cognitive effort.

To test this alternative mechanism, we designed Experiment 4 to make accuracy maximally elastic in effort: the pattern recognition task is self-timed and participants can arrive at a correct answer through a laborious counting exercise. Here, we find evidence that incentives reduce wishful thinking, precisely when incentives lead to increased accuracy. In particular, when we focus on about 40 percent of participants who count more often in response to higher accuracy incentives, we find that these participants, but not others, actually increase their average accuracy and reduce their wishful thinking. These results suggest that material payoffs affect wishful thinking only when they lead to investment and subsequent improvement in signal precision, but not by affecting the inner calculus of ex-post signal distortion. This insight adds nuance to a small but growing literature on the role of accuracy incentives in disciplining motivated beliefs, which, as we discuss in the next section, has yielded several null results.

A final set of results concerns wishful thinking as a personal characteristic or trait. 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 variable across them, with some individuals displaying the opposite tendency. Moreover, an individual's wishful thinking correlates negatively with their self-reported concentration on the pattern recognition task.[2] This is consistent with the idea that ex-ante investments in signal precision decrease wishful thinking. Another intuitive finding is that wishful thinking correlates negatively with a measure of defensive pessimism, a tendency to expect the worst in order to

[2]Therefore, the "natural" differences in concentration between individuals appear to be more consequential than the within-subject shifts in concentration that a higher accuracy bonus managed to induce in the first three experiments.

avoid disappointment. Instead, the correlation between experienced anxiety and wishful thinking is positive. We also find suggestive evidence for a positive correlation between wishful thinking and risk seeking as well as between wishful thinking and the belief in an afterlife, but no correlation with worries about climate change.

In the next section, we review the literature on optimism and wishful thinking. We then describe our experimental design. Section 4 introduces a simple theoretical model that helps us derive our hypotheses. Section 5 contains the results of our experiment before section 6 investigates heterogeneity in wishful thinking. We conclude in section 7.

2 Literature

The literature on wishful thinking, optimism bias, and desirability bias has yielded mixed results and there is still an active debate about the phenomenon's existence, scope and its underlying mechanisms. One strand of this literature uncovers optimism in probability judgements about real life events (Weinstein, 1982; Lench and Ditto, 2008). Other studies claim behavioral and neural evidence for asymmetric updating about future life events, whereby bad news is downweighted (Sharot et al., 2011, 2012). However, these results have been called into question, with critics suggesting that they can be explained by standard Bayesian updating (Shah et al., 2016; Burton et al., 2022). Economic experiments have used more stylized information structures that can rule out Bayesian updating patterns Möbius et al. (2014). However, experiments that study updating from ego-relevant information, such as scores on an IQ test, have yielded mixed results (see Drobner (2022) for a review).

Psychologists have studied optimism using the "marked-card" paradigm, wherein participants rate the probability of drawing a particular card that is associated with monetary outcomes. In a meta-analysis, (Krizan and Windschitl, 2007) find evidence for optimism or desirability bias in this paradigm, but not in other, related settings (see also Windschitl et al., 2010). Some papers on "motivated perception" 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 cannot incentivize beliefs because there is no true state of the world and instead rely on implicit questionnaire items, eye tracking and reaction times (Dunning and Balcetis, 2013) to get at deeply held beliefs. These studies also struggle to rule out that participants believe that their answers can affect outcomes. 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.

In economics there is a small number of studies on wishful thinking in the domain of monetary gains, again yielding mixed results. Mijović-Prelec and Prelec (2010), Coutts (2019), and Mayraz (2011) document overoptimism in estimation tasks in which participants have an exogenous monetary stake in some of the outcomes. However, Coutts (2019) finds it only for one out of three tasks. Barron (2021) finds no evidence for asymmetries in updating of beliefs about the probability of winning monetary prizes.

Our paper differs from previous laboratory studies in several ways. Most importantly, we focus on anxiety as a driver of wishful thinking and on negative rather than positive outcomes. The anticipation of negative events may have a more powerful influence on belief formation both because it may activate different cognitive processes than the anticipation of gains and because people tend to care more about losses than about equivalent gains.[3] This may explain why wishful thinking emerges as a robust phenomenon across our experiments, tasks and contexts, which contrasts sharply with the mixed results in the literature.

Our design also introduces a number of other innovations. First, electric shocks are a proven way of inducing anxiety and allow precise control over its timing. Second, we vary the ambiguity of evidence in a subtle and inconspicuous way. Third, we administer within-subject treatments with many observations per person, which allows us to look at wishful thinking as an individual characteristic that may correlated with other traits. Earlier work in related settings is scarce and has not yielded consistent results. Buser et al. (2018) do not find significant correlations between asymmetric updating of ego-relevant news across three tasks. However, there are few observations per participant and the repeated updating task is noisy and subject to other biases like conservatism and baserate neglect. Sharot et al. (2007) and (Sharot et al., 2012) show neural and hormonal substrates of optimism bias, suggesting a hardwired component that may differ between individuals.

Like us, previous experiments have investigated whether accuracy incentives reduce motivated beliefs, in line with a trade-off between the psychological motive for and the (material) costs of adopting wrong beliefs that is central to models of motivated beliefs (e.g. Bénabou and Tirole 2002; Brunnermeier and Parker 2005; Bénabou and Tirole 2011). In the affirmative, Armor and Sackett (2006) find more optimism for hypothetical than real events and Zimmermann (2020) shows that incentives can reduce motivated biases in recall. However, much evidence goes in the other direction. Simmons and Massey (2012) show that accuracy incentives of up to $50 do not correct football fans' overoptimistic expectations about their home team. Lench and Ditto (2008) find

[3] Falk and Zimmermann (2016) study the role of anxiety in information preferences and investigate whether participants have a preference over early or late resolution of uncertainty about the occurrence of an electric shock.

no effect of incentives on optimistic beliefs about adverse life events. Mayraz (2011) and Coutts (2019) find that higher rewards for accuracy do not reduce wishful thinking, and Schwardmann et al. (2022) finds no evidence for an effect on self-persuasion and polarization in a debating context.[4]

Our results provide a new and more nuanced view of the role of incentives: We find that accuracy incentives only reduce motivated beliefs in tasks where participants can improve the precision of signals trough effort and thereby reduce the scope for wishful thinking. This suggests that the impact of economic incentives on motivated beliefs is likely to be highly sensitive to the nature of the inference task and the extent to which accuracy is elastic in effort.

3 Design

Our study comprises four computerized experiments, which we number in the order in which they were run. We preregistered hypotheses for each experiment on `Aspredicted.org`. The preregistrations and IRB approvals can be found in Appendix C.4.2. All experimental instructions are available in Appendix E.

3.1 Design features common to all experiments

All experiments share the same basic structure. 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 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 pre-

[4]More generally, (Enke et al., 2021) investigate a number of well-known cognitive biases and show that paying up to monthly salary for accuracy does not improve performance, although it does induce more cognitive effort. This confirms an earlier review that concluded that "no replicated study has made rationality violations disappear purely by raising incentives" (Camerer and Hogarth, 1999).

vious literature, we will refer to the emotions the uncertain shock or loss induces as "anticipatory anxiety".[5]

Our first and 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 assured that any differential response to the two types of patterns could not affect our results. The occurrence of the shock depended only on the pre-determined shock pattern and the actual pattern on the screen and not on the participant's response. However, believing that one saw a no-shock pattern could reduce anxiety about the imminent shock or loss. This would bias participants answers towards no-shock patterns, and hence make them less accurate when a shock pattern is shown and more accurate when a no-shock pattern is shown. We measure wishful thinking as the difference between average accuracy for "no-shock" and "shock" patterns.

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

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

3.2 Experiment 1: Electric Shocks

The experiment took place in the CREED experimental laboratory at the University of Amsterdam. Sixty subjects 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

[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	409
Number of trials	216	up to 96	up to 64	up to 96
Visual task	Single Gabor flash	Single Gabor flash	8 Gabor flashes	Colored dots
Anxiety source	Electric shock	Monetary loss	Monetary loss	Monetary loss
Loss/shock size	Self-calibrated	0, 0.1 or 5 pounds	1 pound	0 or 1 pound
Task difficulty levels	Tilt size (3 levels)	Tilt size (2 levels)	Likelihood ratio (continuous)	Dot ratio (4 levels)
Accuracy bonus levels	1 euro 20 euro	10 cents 8 pounds	5 cents 10 pounds	5 cents 10 pounds
Other design elements	Confidence measure Replication exp.		Treatment reminders	Treatment reminders Self-timed task
Start / end date	November 12, 2018 December 5, 2018	Feb 23, 2021 Feb 29, 2021	Jan 3, 2022 Jan 4, 2022	March 8, 2022 March 8, 2022
Location	Laboratory (Amsterdam)	Online (Prolific.co)	Online (Prolific.co)	Online (Prolific.co)

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 (Grillon, 2008; Schmitz and Grillon, 2012; Engelmann et al., 2015, 2019).[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]

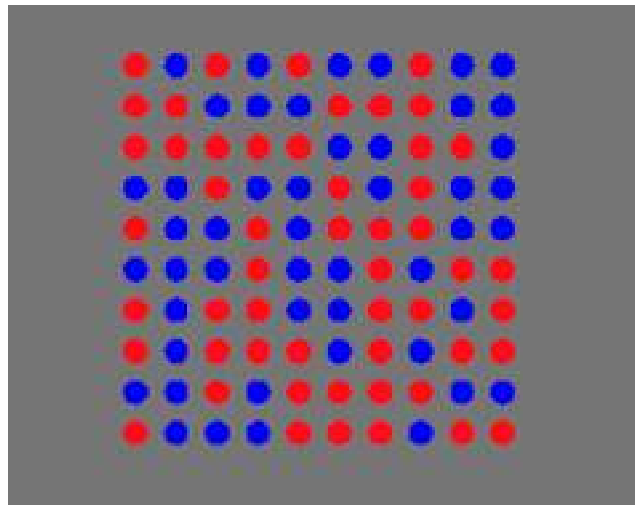
The visual task was to determine whether a grating (Gabor patch), was tilted towards the left or right (see example in Panel (a) of Figure 1). Before each trial, subjects were reminded of the treatment conditions. After briefly seeing a fixation cross (750ms), the grating was flashed on the

[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 (Engelmann et al., 2015, 2019).

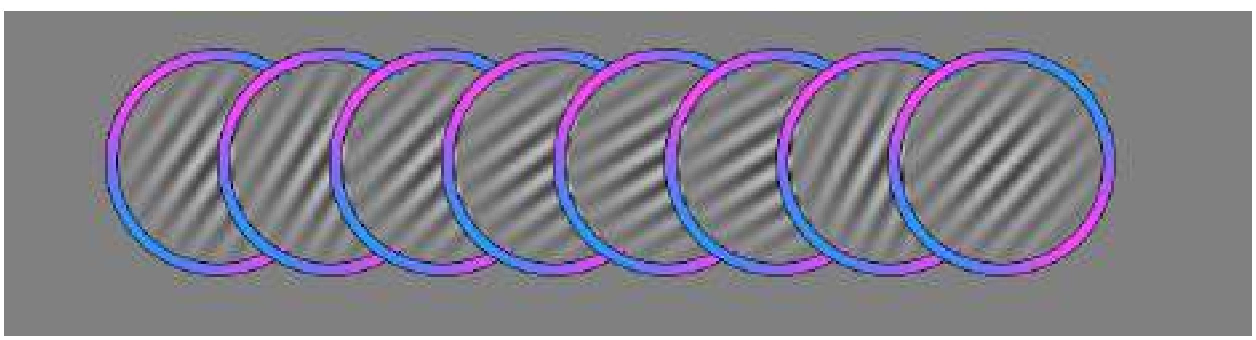
[7]The wrist of the participant's non-dominant hand was connected to a Digitimer DS5 isolated bipolar current stimulator, which itself was connected to MATLAB through National Instruments USB x-series. The participant induced herself with a series of shocks, which she rated on a pain scale of 0 (not painful at all) to 10 (extremely painful). The calibration was complete when the subject rated the pain as 7-9 on the scale three consecutive times. A rating of 10 would lead to a decrease in the threshold. The maximum possible shock strength was set to 5V 25mA and the duration of the shock was set to 50ms (Engelmann et al., 2015, 2019).

screen (150ms). Participants were then asked to indicate the direction of the tilt by pressing the left or right arrow on the keyboard (self-paced) as well as the confidence in their choice on a scale from 50% (completely uncertain) to 100% (certainty). We incentivized confidence ratings with a Becker-deGroot-Marschak (BDM) or "matching probabilities" mechanism. This mechanism makes it incentive compatible to state true beliefs, regardless of a particpant's risk preferences.[8]

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



(b) Dot-counting task (Experiment 4)



(c) Multiple gabor task (Experiment 3)

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

Next, participants faced an anticipation screen (2000-8000ms), asking them to wait for the shock resolution. Finally, the electric shock was administered or not (1000ms). No trial-by-trial feedback was given about the correctness of the guess, but the average performance was communicated at

[8]Subjects indicate their subjective probability $x \in \{50, 55, ..., 95, 100\}$ that their answer was correct. The computer then randomly draws a number $z \in [50, 100]$. If $x \geq z$, then subjects win prize M, if their answer truly is correct. If $x < z$, then subjects win prize M with probability z. M varies between experimental conditions. Schlag et al. (2015) provide details about the origins and incentive compatibility of this mechanism. See Trautmann and van de Kuilen (2014) and Hollard et al. (2016) for evidence. After the instructions but before the experiment started, participants had the opportunity to gain experience with the BDM mechanism.

the end of each block of 18 trials. Each participant completed three sessions, each divided in four blocks of 18 trials. The four blocks correspond to four conditions of a 2x2 factorial design (Shock x Incentive). As described above, the Shock treatment varied whether the possibility of a shock was associated with a right-tilted or left-tilted grating pattern. The Incentive treatment varied whether the potential prize in the belief elicitation was 1 or 20 euros. We also varied the difficulty of the pattern recognition task within each block, as measured by the degree of the tilt from the vertical line.[9]

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

3.3 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 Appendix E). All monetary amounts were communicated in pounds.

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

[9]The three difficulty levels were calibrated to result in accuracy levels of 60%, 70% and 80%. Initially, these levels were calibrated on the basis of a pilot, and were the same for all subjects. To reduce the effects of fatigue or learnings, 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.

[10]As we discuss in Section 2, the connection between monetary outcomes and optimism has previously been investigated by other papers in the positive domain, e.g. Mayraz (2011); Barron (2016); Coutts (2019), which has led to mixed findings.

Using monetary losses allows us to vary the size of the losses, and possibly the anxiety associated with these losses. To this end, each participant went through three parts of the experiment that varied in endowment and loss size: 25 pound endowment with 5 pound losses (*High Loss* condition), a 50 cents endowment with 10 cents losses (*Low Loss* condition), and no endowment with no threat of losses (*Neutral* condition). The part without losses served to address potential confounds that we discuss in Section 5.5.3.

To vary task difficulty, we used two different angles for the tilt of the pattern (with the steeper one being closer to the vertical line and hence harder to recognize). The accuracy incentives varied between trials to be either 8 pounds sterling or 10 cents. Unlike in the previous experiment, we did not elicit confidence measures. Instead, we randomly selected one 8 pound trial and one 10 cent trial and paid subjects if their answer was correct. We made this change to implement the most parsimonious design that still allows for our various treatment dimensions, while avoiding attrition, fatigue and confusion of online participants due to the time-consuming and involved instructions of the confidence elicitations.

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

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

Next to 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, we should ascertain that it is 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 provide enough levers 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. 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.[11]

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 consecutive sequence of 8 Gabor patches spaced over 4 seconds, as illustrated in Figure 1. The tilts were generated from one of two distributions of patterns that was biased towards 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.[12]

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 losses followed that of Experiment 2. Participants completed two parts. In each part they received an endowment of 5 pounds from which they would lose 1 pound with a probability of one third if a "loss pattern" appeared. The part finished when the endowment was exhausted (after 5 losses) or after 32 trials. Within each part of the experiment, there were up to four 8-trial blocks across which we varied the size of the accuracy bonus (0.05 cents vs. 10 pounds) and the orientation of the loss patterns (left vs. right). After each block, there was an interblock survey about concentration on the task. 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 is self-timed, with a time limit of 40 seconds. This allows 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

[11] 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."

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

the screen. Perhaps for that reason, previous studies using these or very similar tasks have found strong effects of incentives for accuracy (Caplin and Dean, 2014; Dean and Neligh, 2019; Dewan and Neligh, 2020). In addition, Bosch-Rosa et al. (2021) find that self-image concerns lead to motivated belief formation in this task.

Other parts of the design followed that of Experiment 2 and 3. Participants completed two parts with 32 trials each. In each part, participants received an endowment of 5 euros from which they would lose 1 euro with a probability of one third if a "loss pattern" appeared. When a subject exhausted the endowment (i.e. after 5 losses), the part stopped. 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 cents vs. 10 pounds) and the color of the loss pattern (blue vs. red). We varied the difficulty of the task, by varying whether the majority color has 51, 52, 53 or 54 dots. In addition, we included one "neutral" part of 32 trials without endowments or losses, the order of which was randomized to be either before or after the two parts with loss trials. Experiment 4 also featured the intertrial self-reports about anxiety and concentration that we used in Experiment 2. For experiment 4, we recruited 409 participants on Prolific.

4 Theoretical predictions

In this section, we present a stylized model of wishful thinking that captures our laboratory 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. The model is in the spirit of Brunnermeier and Parker (2005). The agent chooses her beliefs, trading of the anticipatory utility benefits of optimism with the material costs stemming from wrong decisions. Moreover, belief distortions come at a cognitive cost as in Bénabou and Tirole (2002) and Bracha and Brown (2012).

The state of the world is given by $r_\theta \in \{0, 1\}$, where $r_\theta = 1$ means that the true pattern is *right-tilted*. A participant observes a pattern 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 s, with $\frac{dp(r_\theta=1,s)}{ds} > 0$ and $\frac{dp(r_\theta=0,s)}{ds} < 0$. In particular, they become more certain when the signal is more precise.

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

M

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

The probability of winning the prize is maximized at $\hat{p} = p$. Therefore, if material payoffs were the only object in the agent's utility function, then she would not self-deceive.[13]

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

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

The parameter $\sigma_z \geq 0$ captures the importance of anticipatory utility concerns, or a participant's 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 shocks (hence, the subscript z) are associated with right-tilted ($r_z = 1$) or left-tilted ($r_z = 0$) patterns in a given trial. In our model, anticipatory utility is linear in beliefs.[14]

The agent will not only experience the disutility of anticipatory anxiety, but also the disutility of actually receiving the shock, which is given by $(r_z\hat{p} + (1 - r_z)(1 - \hat{p}))qZ$.

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

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

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

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

[13]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.

[14]Some authors have assumed that anticipatory utility is concave in beliefs (e.g. Caplin and Leahy 2001), which implies that anxiety is then also induced by a greater variance in beliefs. The effect of non-linear belief-based utility on wishful thinking may be a fruitful topic of exploration for future work. What matters for our simple predictions here is that utility is monotonic in beliefs, we do not require linearity.

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

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

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

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

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

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

The 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 impacts the motivation to hold distorted beliefs. But because, owing to our symmetric design, $p(s, r_\theta)$ drops out of our measure of wishful thinking, we can shed the light on 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 declines in the size of the accuracy bonus, i.e. $\frac{dW}{dM} < 0$.*

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

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

Appendix C features a number of extensions of the model. First, in Section C.1, we show that the above predictions are robust to allowing the agent to derive anticipatory utility from her expectation of future accuracy payoffs. We show that the agent's wishful thinking increases in this case, because her beliefs will be less disciplined by accuracy incentives. She now cares about the actual *and* the perceived accuracy payoffs she obtains and the latter are not decreasing in biased beliefs.

Second, in Section C.2, we allow for a “bracing” or “defensive pessimism” motive for biased beliefs. 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. We show that 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.

We can also use the model to account for correlations between measures of wishful thinking and (realized) anxiety, based on the interpersonal heterogeneity of fundamental parameters. In Section C.3 we show that this correlation can be positive, negative or zero depending on which underlying heterogeneity is driving differences in wishful thinking and anxiety. In particular, heterogeneity in λ implies a negative correlation and heterogeneity in σ_z implies a positive correlation. The empirical correlation therefore sheds light on the relative importance of different heterogeneities.

Our data confirms some predictions of the model and is at odds with some others. To capture these discrepancies, section C.4 proposes a revised model that allows for ex-ante investments in signal precision.

5 Results

We start with an overview of the main results across all of our experiments. The outcome of interest is whether pattern detection accuracy differs between shock/loss and no-shock/no-loss patterns. Table 2 shows the results of OLS models that regress 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. The Appendix contains further data and analyses. In Appendix A we show that all of our results are robust to panel data regressions over all trials that include individual fixed effects and cluster standard errors at the individual level. Appendix Table A.1 provides descriptive statistics of accuracy levels for all of our experiments. Appendix Figure **??** provides an overview of the cumulative distribution functions of accuracy in shock/loss and no-shock/no-loss patterns.

Our main hypothesis is that participants are less accurate in identifying patterns associated with a shock or monetary loss. Table 2 exhibits strong evidence for this hypothesis. In all experiments, the coefficient for shock/loss patterns is negative and highly statistically significant. This also holds when we include interactions with the Difficulty and High Accuracy Bonus treatments.

The only qualification to this statement is needed in Experiment 1 (column 2), where we find statistical significance only for the more difficult patterns. Thus, wishful thinking appears as a robust phenomenon, both across sources of anxiety and across pattern recognition tasks.

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 difficultly level across patterns shows participants are less likely to be correct on difficult patterns, where the size of the coefficient reflects how we operationalized difficulty in the different experiments (see Table 1 for details). Moreover, the interaction term shows that this is especially pronounced for loss or shock patterns, thus confirming our hypothesis in all experiments.

The third hypothesis is that incentives for accuracy reduce wishful thinking, as 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. However, a closer examination, in Section 5.3, reveals that accuracy incentives can have an effect in some settings. Finally, we find that loss size, which we varied in Experiment 2 has a positive effect on accuracy and increases wishful thinking, but in both cases the effects are not statistically significant.

In the sections below, we elaborate on the results of the individual experiments and develop additional insights and interpretations.

As Table 2 shows, effect sizes differ quite a lot across experiments. For instance, effect sizes are about four times larger in Experiment 2 than in Experiment 1. It is difficult to directly compare these effects, as there were several differences between these experiments. Besides replacing shocks with losses, Experiment 2 took place online, which necessitated changes to the exact instructions, the earning amounts and number of trials. Experiments 3 and 4 further differ in the perceptual task and other implementation details.

If we average wishful thinking over all participants in all experiments, then we find that average accuracy is 78.0 percent for no-loss/shock patterns and 69.7 percent for loss/shock patterns, respectively 28.0 percentage points and 19.7 percentage points above the 50 percent baseline of random choice.[15] Therefore, average wishful thinking is 8.3 percentage points and seeing a shock/loss rather than a no-shock/no-loss pattern decreases performance above chance by almost one third.

[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.

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

	Experiment 1 (Electric Shocks)		Experiment 2 (Monetary losses)		Experiment 3 (Repeat flash)		Experiment 4 (Dot task)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep var:	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy	Accuracy
Shock/Loss pattern	-4.111***	-2.014	-16.54***	-8.248**	-4.266***	-3.052***	-8.453***	-7.393***
	(1.264)	(1.736)	(1.605)	(3.489)	(0.766)	(0.865)	(1.040)	(1.308)
High accuracy bonus (HAB)	0.785	0.313	-0.588	-1.081	0.630	0.685	1.670***	1.004
	(0.878)	(1.387)	(0.851)	(1.089)	(0.474)	(0.601)	(0.627)	(0.853)
Difficulty level (DL)	-8.602***	-7.318***	-15.68***	-11.04***	-20.55***	-19.39***	-7.073***	-6.497***
	(0.634)	(0.795)	(1.019)	(1.114)	(0.668)	(0.794)	(0.269)	(0.361)
Loss size (LS)			-0.617	0.776				
			(0.906)	(1.245)				
Shock/Loss pattern x HAB		0.944		0.994		-0.110		1.330
		(1.787)		(1.771)		(0.881)		(1.322)
Shock/Loss pattern x DL		-2.569**		-9.200***		-2.317***		-1.150**
		(1.102)		(1.701)		(0.892)		(0.503)
Shock/Loss pattern x LS				-2.784				
				(1.869)				
Constant	80.75***	79.70***	85.82***	81.65***	87.66***	87.06***	89.53***	89.00***
	(1.106)	(1.287)	(1.964)	(2.310)	(0.791)	(0.829)	(0.732)	(0.796)
Observations	720	720	3415	3415	3408	3408	6534	6534
R^2	0.261	0.266	0.134	0.140	0.236	0.236	0.110	0.110

OLS regressions of accuracy on treatment dummies and interactions. Each observation is the average accuracy of an individual over all trials in a given treatment. "Shock/Loss pattern" is a dummy if the pattern is associated with a shock (Experiment 1) or loss (Experiments 2-4). "High accuracy bonus" is a dummy that represents a high accuracy bonus, while "Difficulty level" is a categorical variable that counts the difficulty level of the perceptual task, with the number of levels dependent on the experiment (see Table 1 for details). The continuous difficulty levels in Experiment 3 were binarized using a median split. "Loss Size" refers to the size of the monetary loss that we varied in Experiment 2. Standard errors in parentheses clustered by individual. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Our effect sizes are unlikely to be predictive of particular applications, since our results show a dependence on the task at hand and its difficulty. Nevertheless, as an external benchmark, one might consider mammogram reading, a complex pattern recognition task with a high-stake emotional outcome. Studies on interventions with radiologists often celebrate improvements in accuracy of a few percentage points, which are well in range of our effect sizes (Hadjiiski et al., 2004; Houssami et al., 2004).

5.1 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.[16] Panel (a) compares average accuracy between shock and no-shock patterns, demonstrating wishful thinking of about four percentage points (72.3 vs. 68.6 percent). Panel (b) displays the impact of the three difficulty levels, as defined by the size of the tilt of the pattern. There appears to be some wishful thinking for easy patterns (2.4 percentage points) and medium patterns (2.5 percentage points). However, Table A.2 provides interaction terms for each of the difficulty levels, and shows that wishful thinking is statistically significantl only for the most difficult patterns, where it rises to about 8 percentage points. Finally, panel (c) displays the impact of raising the prize for the BDM mechnism from 1 to 20 euro. Wishful thinking is about 1.4 percentage points more pronounced under the low bonus than under the high bonus, but the difference between the two conditions is not statistically significant.

Confidence. In addition to the accuracy measure, we elicited a measure of confidence in having correctly identified the pattern, incentivized with a BDM mechanism. Since we know the actual state of the world, this allows us to construct the variable "Belief", which measures the subjective belief in the correct answer and provides a more continuous measure of participants' perceptions. Beliefs vary on a scale from 0 (meaning the subject indicated 100% confidence in the wrong answer) to 100 (meaning the subject indicated 100% confidence in the correct answer). Figure A.1 and Table A.3 in 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.

[16]Table 2, in columns 1 and 2, provides the regression analysis associated with these results, and Appendix Table A.2 adds robustness across regression specifications. Moreover, in Appendix B, we provide a replication of Experiment 1 with $N = 50$.

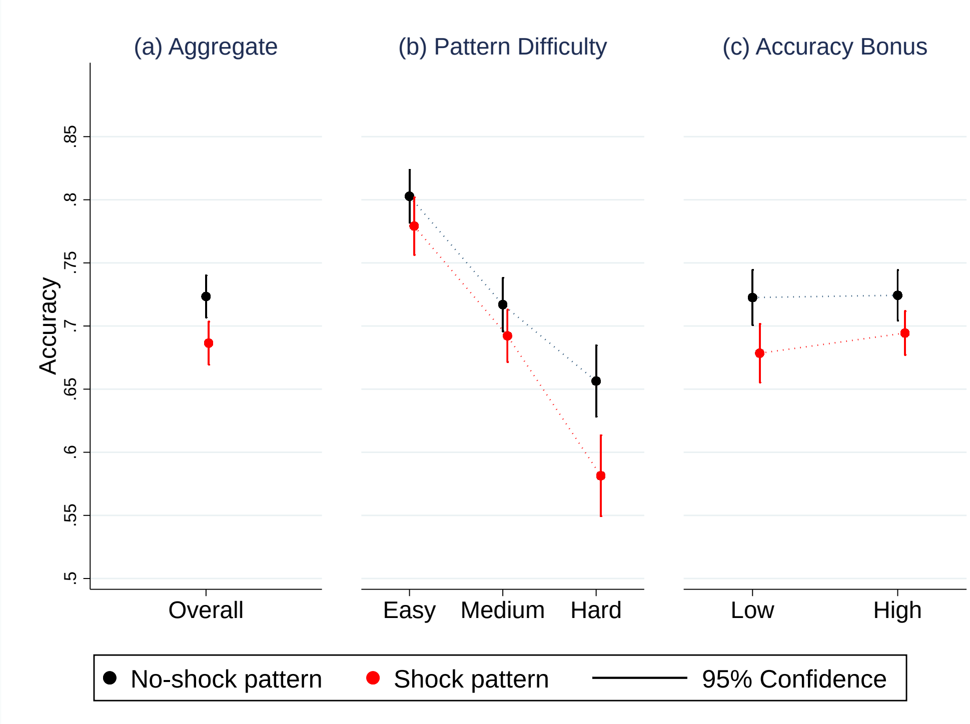


Figure 2: Electric shocks and accuracy in Experiment 1. Average accuracy levels, split by shock and no-shock pattern. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category, so $N = 60$ in each category. Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty (tilt) of the pattern. Panel c) disaggregates by incentives for accuracy.

5.2 Experiment 2: Monetary losses as a source of anxiety

We now turn to Experiment 2, which replaced electric shocks with monetary losses. While the literature has documented how the threat of electric shocks increases anxiety, no such evidence is available for losses. As a manipulation check, we therefore asked subjects to report their agreement with the statement "I felt anxious about losing money from my endowment" on a scale from 1 to 5 after each treatment block of 8 trials in which losses could occur. Panel (a) of Figure 3 shows the outcome of this manipulation check, where we count the scores in each block in both the Low Loss (10 cents) and High Loss (5 pounds) condition. This shows a clear difference between the two loss conditions, with average anxiety being 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). It also demonstrates 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, provides regression evidence

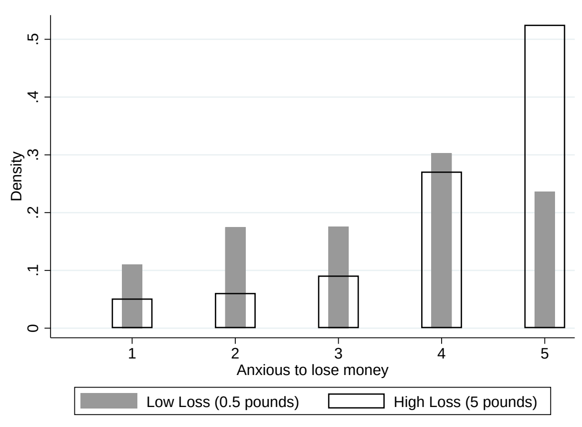


Figure 3: Manipulation check. Histogram of agreement with the statement "I felt anxious about losing money from my endowment" measured on a 5-point Likert scale, split by loss size. Each report in a treatment block counts as one observation.

associated with these results, and Table A.4 provides robustness across regression models. Results exclude the Neutral condition, since this is not a test of wishful thinking and is discussed in Section 5.5.3.

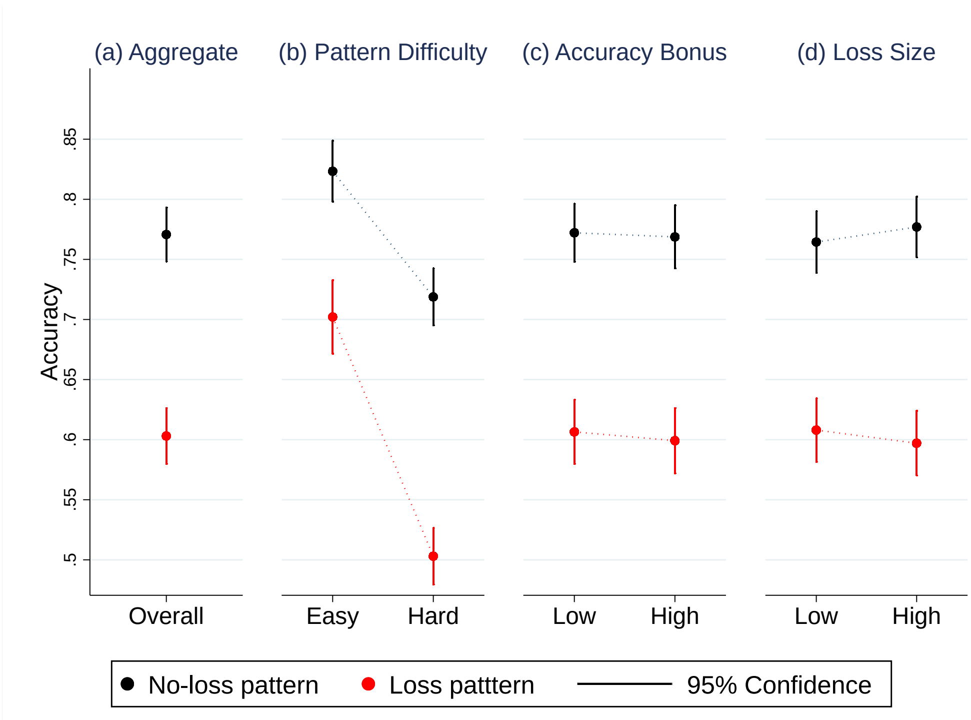


Figure 4: Monetary losses and accuracy in Experiment 2. Average accuracy levels, split by loss and no-loss patterns. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category, so $N = 221$ in each category. Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty (tilt) of the pattern. Panel c) disaggregates by incentives for accuracy. Panel d) disaggregates by size of losses.

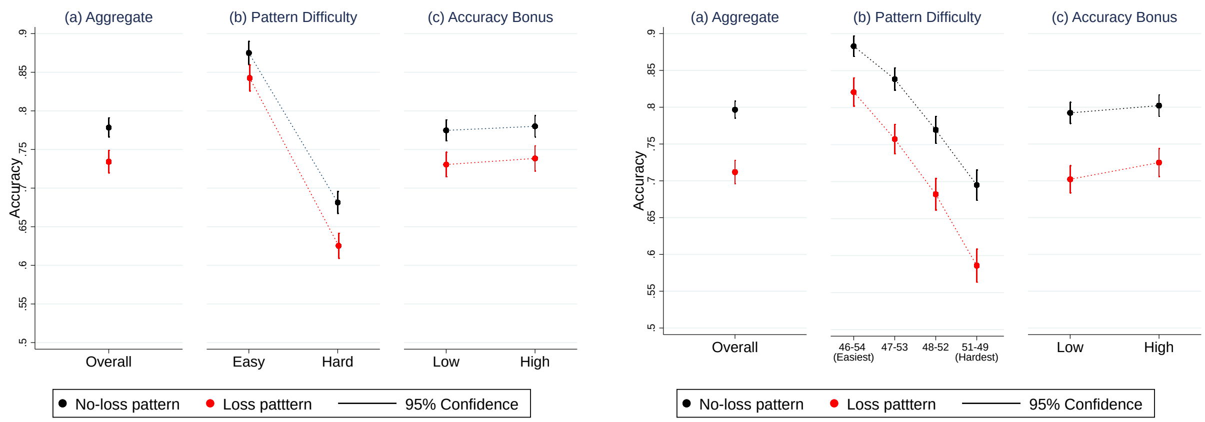
Panel (a) of Figure 4 compares average accuracy on No-loss patterns with the Loss patterns. We see wishful thinking of 17 percentage points, which is highly statistically significant. The effect size is strikingly large: compared to the random-choice benchmark of 50 percent accuracy, the improvement in accuracy is almost 3 times higher under patterns associated with no loss compared to those that are associated with a loss. Panel (b) shows clear wishful thinking for both pattern difficulty levels, as well as an interaction effect between wishful thinking and difficulty. Panel (c) shows the effect of seeing a loss pattern for accuracy bonuses of 0.1 and 8 pounds respectively. We find a clear evidence for wishful thinking in both cases. However, there is no evidence that incentives improve performance, and no evidence for an effect of the accuracy bonus on wishful thinking, with wishful thinking being 16.6 percentage points under the low bonus and 17.0 percentage points under the high bonus. We will come back to the interpretation of this null result in Section 5.4. Finally, Panel (d) shows the effect of changing the loss size from 10 cents to 5 pounds. While this raises wishful thinking by about 2.7 percentage points, this difference is not statistically significant.

5.3 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. Columns 4-8 of Table 2 provide regression evidence associated with these results and Tables A.5 and A.6 provide evidence for robustness across regression models.

Results Experiment 3. Subfigure (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. This results in wishful thinking of 4.4 percentage points. Task difficulty was a continuous variable in this task, defined by the posterior likelihood ratio of the two pattern-generating processes. Figure 5 Panel (b) displays the impact of a median split on this variable, and shows both a clear and statistically significant effect of higher difficulty on wishful thinking. Panel (c) shows wishful thinking for accuracy bonuses of 0.05 and 10 pounds respectively. Again, we find little evidence that incentives improve performance: A high bonus improves accuracy by about 0.7 ppt, but the effect is not close to being statistically significance. Moreover, there is no interaction with the loss pattern, so no reduction in wishful thinking from higher accuracy incentives.

Results Experiment 4. Subfigure (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



(i) Sequential gabor task (Exp. 3)

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

Figure 5: Accuracy in the multiple gabor and dot-counting tasks in Experiment 3 and 4. Average accuracy levels, split by loss and no-loss pattern. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category. In each subfigure, Panel a) shows aggregate results. Panel b) disaggregates the results by difficulty of the pattern, with a median split shown for Experiment 3. Panel c) disaggregates by incentives for accuracy.

in this task. Panel (b) displays the impact of pattern difficulty, where the easy patterns had a 46-54 split in colored dots, and the hardest patterns a 49-51 split. Once again, we confirm a statistically significant effect of difficulty on accuracy as well an interaction with wishful thinking. Panel (c) shows the pattern for the different levels of the accuracy bonus of 0.05 and 10 pounds. Unlike for the tasks we considered above, incentives improve performance: A high bonus improves accuracy by about 1.6 ppt, which is significant at the 5 percent level in the regression in Table A.1. However, as in our previous experiments, there is little interaction with the loss pattern. As we discuss next, this result hides important heterogeneities.

5.4 Incentives for accuracy

Across our experiments, we find no statistically significant effect of the accuracy bonus on wishful thinking. This may indicate that subjects do not make cognitive trade-offs between psychological and material incentives. Alternatively, it may be because participants did not care about or pay attention to the accuracy bonus or somehow were not able to react to it. We now discuss these explanations in more detail, and find that our aggregate results hide some important heterogeneity.

Incentives and cognitive effort. To investigate whether participants cared about, paid attention to, and responded to the bonus, we asked participants in Experiments 2, 3 and 4, at the

Table 3: Regressions of cognitive effort on accuracy bonus across experiments

	(1) Concentration	(2) Concentration	(3) Concentration	(4) Response time (log)	(5) Response time (log)	(6) Response time (log)	(7) Response time (log)
High acc. bonus	0.117***	0.120***	0.172***	0.0377**	0.0525***	0.0311***	0.130***
	(0.0330)	(0.0169)	(0.0258)	(0.0169)	(0.0129)	(0.00646)	(0.0176)
Experiment no.	2	3	4	1	2	3	4
Fixed effects				✓	✓	✓	✓
Observations	442	852	818	11520	11396	33507	21217
R^2	0.007	0.017	0.012	0.001	0.003	0.001	0.013

Regressions of cognitive efforts on a dummy for the high accuracy bonus by experiment. Columns 1-3 show regressions on self-reported concentration, where an observation is an individual's average concentration over all trials in the High Bonus and Low Bonus condition. Concentration is measured as agreement with the statement "In the past 8 trials I was very concentrated on the task" on 5-point Likert scale. Columns 4-7 show panel regressions where the outcome variable is log response time in each trial, where the latter is measured in milliseconds. Panel regressions include individual fixed effects. Standard errors in parentheses are clustered by individual. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

end of each 8 trial block, to report their agreement with the statement "I was very concentrated on the task". Columns 1-3 of Table 3 show regressions that identify the increase in this variable resulting from a higher accuracy bonus, where each observation is an individual's average reported concentration for each accuracy bonus level. In each of the experiments, a higher accuracy bonus leads to a significant increase in self-reported concentration.

In addition, if the high bonus leads to a more careful evaluation of answers, one would expect longer response times. Because of the highly the skewed nature of the response time distribution, which may be sensitive to large outliers, we look at the logarithm of response times as an outcome variable, which is measured in milliseconds. Columns 4-7 of Table 3 show panel regressions where we regress log response times on the high accuracy bonus, revealing an increase in response times. The accuracy bonus significantly increases response times in all experiments (we find similar results if we take raw response times), particularly in Experiment 4, which we will investigate in more detail below.

These results imply that participants care about and react to the accuracy bonus, albeit not by adjusting their wishful thinking. We can look at experiment 2, which featured the most wishful thinking of all experiments, to calculate an upper bound of the monetary cost associated with this stubborn wishful thinking. We zoom in on trials in which the loss and the correct answer were aligned, which mirror the many applications where what is true and what scares us aligns.[17]

[17]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

Comparing accuracy on such loss patterns in the High Bonus condition with accuracy in a set of control trials in which no losses were possible, implies an expected monetary cost from wishful thinking of about 87 cents.[18] This corresponds to roughly 10 minutes of work on the Prolific platform.

Incentive effects and cognitive control. The only time we find a statistically significant effect of accuracy incentives on accuracy is in the context of the dot counting task in Experiment 4. This experiment also sees the largest increases in our measures of cognitive effort in Table 3. This makes sense, as this task was chosen to be very elastic to cognitive effort. Participants who are really motivated to get the answer right may even *count* the dots. Incentive effects on performance and wishful thinking may therefore be particularly pronounced if they motivate participants to engage in such a strategy.

To explore this, we asked participants in the post-experimental questionnaire whether they counted dots. On this question 9% of subjects replied with "Always", 38% replied "Sometimes", and 53% replied "Never". These answer categories do indeed 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, the participants in the Sometimes category also show a 32 percent increase in mean response times in reaction to the high accuracy bonus. This effect is far larger than that observed for participants in the Never (7 percent) and Always (14 percent) categories. Given the highly skewed nature of response time distributions appendix Table A.8 also analyzes the effect of the bonus on median response times. This shows that the increase is not significant for the Never category, but highly significant in the Sometimes category ($p < 0.0001$). Moreover, it is marginally significant for the Always category and highly significant if we combine the Sometimes and Always categories.

Since the Sometimes counters are most responsive to the accuracy bonus, we investigate whether this category is also more likely to exhibit a reduction in wishful thinking. Figure 6 shows wishful thinking across the three counting categories. The figure shows clear evidence for a reduction in

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

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

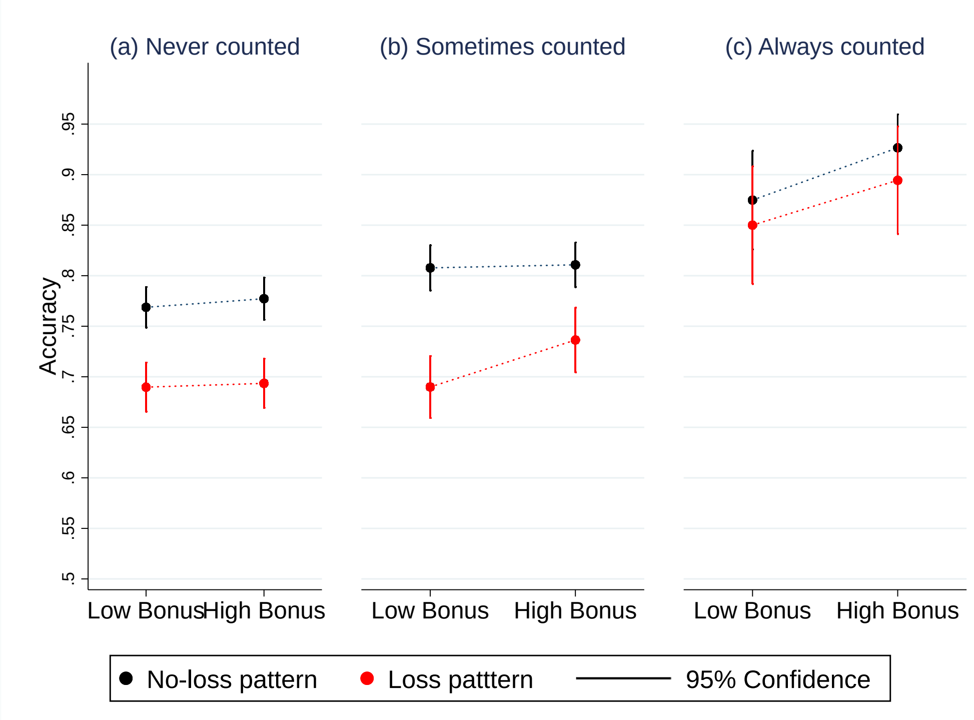


Figure 6: Accuracy in the dot-counting task. Average accuracy levels, split by loss and no-loss patterns. Bars indicate 95% confidence intervals. One observation is the average over an individual's trials in a given category. Panel a) shows participants who report that they never count ($N = 214$). Panel b) shows participants who sometimes counted ($N = 154$). Panel c) shows participants who always counted ($N = 35$).

wishful thinking among the Sometimes category. Appendix Table A.7 shows that this interaction is significant at the 5 percent level for that category, and marginally significant for all counters. The Never category show a slightly negative and insignificant interaction. The variation in accuracy between the categories of counters provides further evidence that effort reduces the scope for wishful thinking: the Always counters find the correct answer substantially more often than other participants and also show a highly reduced, and statistically not significant level of wishful thinking.

The results in this section provide evidence that incentives for accuracy can reduce wishful thinking. However, such an effect only obtains for participants who increase their accuracy through cognitive effort. The increase in accuracy may then act much like the exogenous decrease in pattern difficulty we saw affecting wishful thinking. By contrast, the self-deception we see taking place in all of our experiments does not seem to respond to material incentives directly, i.e. by participants calibrating the intensity of their wishful thinking or ex-post signal distortion to trade-off psychological and material incentives at the margin.

How do these results relate to our model in Section 4? The model predicts that a higher accuracy bonus reduces wishful thinking, by affecting the optimization process behind motivated beliefs. The prediction is consistent with the behavior of counters in Experiment 4. Thus, in settings that allow for successful investments in signal precision, the model predicts correctly, even if only in an "as-if" sense. However, the model predictions fail in the setting of Experiments 1, 2 and 3, which precluded the possibility of improving accuracy, suggesting that accuracy incentives do not affect the ex-post signal distortion featured in the model.

In Appendix section C.4, we revise the model to be more in line with our findings. There, we assume that self-deceptive efforts are costless up to a certain point, but impossible thereafter. One interpretation is that self-deception is closer to an "automatic" or "system 1" process (see also Kappes and Sharot, 2019). This model implies that wishful thinking is slow to respond to psychological and material incentives at the margin. Successful investments in signal precision, on the other hand, can constrain wishful thinking by lowering the maximum possible amount of self-deception. Higher accuracy bonuses increase such investments, irrespective of whether agents are sophisticated about the effect of signal precision on wishful thinking, and, in line with our data, then reduce wishful thinking whenever they are successful.

5.5 Robustness and competing explanations

In this section we conduct a number of robustness checks and investigate competing explanations for our results.

5.5.1 Robustness

At the end of the experiment, we asked subjects several questions about their perceptions of the experiment. We use these variables to conduct several robustness checks that could identify the potential effects of confusion, misunderstanding, or distrust in the experimenter on wishful thinking. In these analyses, we pool the data from all experiments. In a first check, we restrict our sample by excluding participants who scored high on perceived difficulty of the instructions, a general measure of understanding. In further robustness checks we exclude those 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 measure excludes some particpants who answer almost randomly and a small number of participants who almost always select the no-shock pattern.

The results are reported in Appendix Table A.9: wishful thinking remains highly significant in all selected samples, with small and statistically insignificant changes in effect sizes. The interaction of shock patterns with pattern difficulty also remains statistically significant in all specifications. The estimate for the interaction effect between the accuracy bonus and the shock pattern is generally positive but not statistically significant. Table A.10 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.

5.5.2 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 non-aligned pattern in order to avoid the shock or loss.

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 77 percent of subjects who correctly gave the first answer had an average wishful thinking of 8.32 percentage points, while those who selected one of the other answers had indistinguishable average wishful thinking of 8.30 percentage points ($p = 0.99$, t-test). In column 7 of Appendix Tables A.9 and A.10, 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.

5.5.3 Does seeing a shock or loss pattern increase noise?

It is possible that seeing a pattern that is associated with a loss or shock increases noise in their answer, 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.

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

To test this prediction, we conducted a Neutral treatment in both Experiment 2 and Experiment 4. In one part of the experiment, implemented in random order, subjects were informed that they could not lose money from their endowment in any trial of this part. We compare accuracy for neutral patterns with accuracy for loss and no-loss patterns, where we pool the data from the two loss sizes in Experiment 2. As before, we take as an observation the individual accuracy rate in each of these conditions. In both Experiment 2 and 4, we find that average accuracy for neutral patterns is between that of the loss and no-loss patterns. In Experiment 2, respective accuracy rates are 71.2 percent (neutral), 60.3 percent (loss), and 77.0 percent (no-loss). In Experiment 4, the corresponding percentages are 75.7 (neutral), 71.1 (loss) and 79.7 (no-loss). All within-experiment differences are statistically significant in a regression analysis (see appendix Table A.11). Furthermore, there is not much evidence that stress impedes average performance: accuracy is slightly (2.7 percentage points) higher in the Neutral condition than the average of the Loss and No-loss condition in Experiment 2, but not in Experiment 4, where they are almost identical (see Appendix Table A.1). Finally, a Neutral treatment in the replication of Experiment 1 further confirm these patterns, details of which are in Appendix B. We conclude that the data reject the noise-based explanation.

5.5.4 Dynamics

Our experiments consist of many trials and within-subject treatments, so we can ask how wishful thinking evolves over time. It may be the case that participants get desensitized to the anxiety-inducing effects of electrical and monetary shocks as they experience an ever greater number of trials. They may then exhibit less wishful thinking in later trials. It may also be the case that initial experiences with losses or shocks heighten subsequent anxiety and increase wishful thinking in later trials. Our dataset offers a window into how motivated beliefs respond to experience and speaks to mechanisms that may be at play in real world settings, which often feature dynamics and an element of repetition.

In Appendix Figure A.5 we provide a visual overview of wishful thinking over time in each experiment. Appendix Table A.12 analyses statistically how the effect of seeing a loss or shock pattern on accuracy (our measure of wishful thinking) evolves over time by interacting a dummy for whether a participant sees a loss pattern with the number of trials a participant has gone through. In a second set of analyses we simply compare wishful thinking in the first half and the second half of the experiment. In Experiment 1, wishful thinking in the first half of the experiment is more than twice as large as wishful thinking in the second half. The coefficient is just shy of statistical significance at conventional levels ($p = 0.102$), but suggestive of the idea that participants get distracted or desensitized to the anxiety-inducing effects of electric shocks as trials go by. On the other hand, 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.

We can also ask how the experience of a shock or loss in the previous trial affects wishful thinking in the current trial. Appendix table A.13 investigates the effect of lagged shocks or losses on wishful thinking. We see no effect in any of the four experiments, regardless of whether or not we control for the time trend in wishful thinking.

While the structure of our dataset allows us to analyse the evolution of wishful thinking with experience, no unambiguous story emerges. The data suggest that repeated shocks may lead to desensitisation, but that experience with monetary losses can lead to heightened wishful thinking in some contexts. This suggests that the source of anxiety may matter for the dynamics of wishful thinking.

6 Wishful thinking as a trait

Motivated cognition, of which wishful thinking is one example, 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, with a few exceptions noted in Section 2, has therefore not been able to say much about individual differences. Instead, our within-subject design with many trials allows us to compute individual measures of wishful thinking and ask two underexplored questions: are there individual differences in wishful thinking and, if so, does wishful thinking correlate with other individual characteristics?

6.1 Wishful thinking as a personal trait

Do individuals differ in their proclivity for wishful thinking? Appendix Figure A.3 depicts histograms of individual-level wishful thinking in each experiment. We see that there is substantial variance, with a majority of participants engaging in some wishful thinking and some participants exhibiting the opposite treatment 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

Table 4: Half-split correlations

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

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

wishful thinking in the other half. For this exercise we split trials into odd and even numbered trials, trials with easy and trials with difficult patterns, trials in the first half and trials in the second half of the experiment and, for Experiment 2, trials with high stakes and trials with low stakes. Calculating such half-split correlations is common in psychology, where they are used to assess the reliability of individual measures derived from cognitive tasks (for example, see Pronk et al. 2021).[19]

Columns 1 through 3 of Table 4 report the half-split correlations of Experiments 2, 3 and 4 respectively.[20] Correlations are around 0.5, with some fluctuations depending on how we split the data, indicating that heterogeneity in wishful thinking reflects individual differences. Moreover, our measure of wishful thinking is only slightly less reliable or stable than participants' skill in the pattern recognition tasks, as shown by the half split correlations in accuracy that we report in columns 4 through 6 of Table 4. To further show that our results are not driven by a few outliers, Appendix Figure A.4 shows the scatterplots pertaining to the odd-even trial splits in Table 4.

6.2 Emotional and cognitive covariates of wishful thinking

Since wishful thinking appears to be a stable individual characteristic, a natural next question to ask is whether it correlates with other variables, and whether such correlations can help understand the drivers of wishful thinking. First, we look at a self-reported measure of concentration, which we measured in the interblock surveys of Experiment 2, 3 an 4. Increased concentration may lead to more precise perceptions and higher accuracy, which in turn constrains wishful thinking. This

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

[20]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.

is the mechanism suggested by our results on dot counters in Experiment 4.

A second covariate of interest is "defensive pessimism", which measures the degree to which people adopt pessimistic beliefs to avoid disappointment.[21] 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).[22] Defensive pessimism runs counter to wishful thinking, as we show formally in Appendix C.2, so one would expect a negative correlation.

Finally, we are interested in the correlation of wishful thinking with self-reported anxiety about losing money from the endowment, which we measured in the interblock survey in Experiments 2 and 4 and hypothesized as a key antecedent of wishful thinking. The sign of this correlation is theoretically ambiguous, as we explain formally in Appendix Section C.3. If participants' vary strongly in their ability to self-deceive, then higher wishful thinking should be associated with lower experienced anxiety. Conversely, if the primary source of heterogeneity between participants' wishful thinking is their proneness to anxiety, then wishful thinking should be positively correlated with experienced anxiety.

Table 5 shows OLS regressions of wishful thinking on these 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 task. Column 1 shows that wishful thinking is negatively correlated with the average self-reported concentration on pattern recognition. Wishful thinking therefore appears to be constrained by cognitive effort, presumably because participants who concentrate more are able to generate significantly more accurate representations of the signal.[23]

The correlation between wishful thinking and defensive pessimism 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, with significance at

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

[22] The presence of both anticipatory utility motives and a desire to avoid disappointment also has implications for the likely time-path of beliefs (Macera, 2014). Unfortunately, we only observe static beliefs.

[23] Regressing participants' accuracy on their self-reported concentration, while controlling for experiment fixed effects, yields a highly statistically significant coefficient ($p < 0.001$): going up one step on the five point Likert scale on which we measured concentration increases accuracy by 2.5 percentage points.

Table 5: Emotional and cognitive covariates of wishful thinking.

Dep. variable:	(1) Wishful Thinking	(2) Wishful Thinking	(3) Wishful Thinking	(4) Wishful Thinking
Concentration	-0.0289***	-0.0345***	-0.0397***	-0.0498***
	(0.00958)	(0.0122)	(0.0107)	(0.0134)
Defensive pessimism	-0.00604	-0.0107*	-0.00905**	-0.0149**
	(0.00399)	(0.00609)	(0.00425)	(0.00678)
Anxiety		0.0157*		0.0198**
		(0.00826)		(0.00891)
Constant	0.328***	0.317***	0.385***	0.386***
	(0.0494)	(0.0640)	(0.0559)	(0.0733)
Experiment dummies	✓	✓	✓	✓
Restrictions	None	None	Difficult instructions <4 of 7	Difficult instructions <4 of 7
Observations	1049	624	743	421
R^2	0.066	0.054	0.086	0.077

OLS regressionsof wishful thinking on cognitive covariates. Data are from experiments 2, 3 and 4 in columns 1 and 3 and from experiments 2 and 4 in columns 2 and 4. Columns 3 and 4 only include participants with one of the three lowest scores on the question "How difficult did you find it to follow the instructions of this experiment?" measured on a 7-point Likert scale from very easy to very difficult. All regressions contain experiment dummies. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

marginal levels.[24] Furthermore, the negative correlation between defensive pessimism and wishful thinking is now also marginally significant. In columns 3 and 4 we do a robustness check on these correlations, akin to that in Section 5.5.1, and exclude participants who reported that they found following the instructions hard to follow. Excluding such potentially noisy participants results in stronger correlations between wishful thinking and all covariates, adding confidence to the results.

These results allow us to sharpen our interpretations of wishful thinking. First, the negative correlation with concentration further underscores how cognitive effort can constrain wishful thinking through its effect on accuracy. Second, the negative correlation with defensive pessimism suggests that belief-based utility motives that run counter to wishful thinking exist and can be detected in the cross-participant heterogeneity of belief biases. Since defensive pessimism is a self-reported survey scale, its correlation with wishful thinking suggests that people are at least somewhat conscious of their tendencies for probability distortion. Finally, the positive correlation with self-reported

[24]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=0.3, $p < 0.001$), thereby validating our more focused and tailor-made measure. However, perhaps not surprisingly, the positive correlation between BAI and wishful thinking is not statistically significant.

Table 6: Real world attitudes and beliefs

Dep. variable:	(1) Risk seeking	(2) After-life	(3) Climate worry	(4) Risk seeking	(5) After-life	(6) Climate worry
Wishful Thinking	0.489*	0.365	0.0295	0.762**	1.018**	0.110
	(0.254)	(0.354)	(0.267)	(0.319)	(0.448)	(0.348)
Constant	3.584***	3.416***	5.704***	3.577***	3.228***	5.675***
	(0.121)	(0.169)	(0.128)	(0.159)	(0.223)	(0.174)
Experiment dummies	✓	✓	✓	✓	✓	✓
Restrictions	None	None	None	Difficult instr. <4 of 7	Difficult instr. <4 of 7	Difficult instr. <4 of 7
Observations	1007	1007	1007	724	724	724
R^2	0.004	0.003	0.018	0.009	0.010	0.025

OLS regressions of wishful thinking on real world beliefs and attitudes. Risk seeking was measured as the answer to the question ”Are you rather a risk-taking or risk-averse person (trying to avoid risks)?” on a Likert scale from 1 (very risk-averse) to 7 (very risk-seeking). Afterlife and climate worry were measured as agreement with the following statements on a 7-point Likert scale: ”I believe in the existence of an afterlife.”, ”I am worried about climate change”. Data is from experiments 2, 3 and 4 in columns 1, 2, 4 and 5. Data is from experiments 2 and 4 in columns 3 and 6. Columns 3 and 4 only include participants with one of the three lowest scores on the question “How difficult did you find it to follow the instructions of this experiment?” measured on a 7-point Likert scale from very easy to very difficult. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

anxiety suggest that people differ in their innate anxiety, and that these differences are not (fully) overcome by their wishful thinking.

6.3 Does wishful thinking explain real world attitudes and beliefs?

To see whether wishful thinking helps explain other attitudes and beliefs, Table 6 features OLS regressions of three unincentivized and self-reported outcomes: risk seeking, belief in an afterlife, and worry about climate change. Columns 1 to 3 include data from all participants from whom the relevant dependent variables were elicited. Columns 4 to 6 again restrict the sample to participants who found the instructions easy to understand.

In column 1, we find a positive and marginally significant correlation between wishful thinking and risk seeking and in column 2, a non-significant positive correlation between wishful thinking and belief in an afterlife. These correlations become more significant in the restricted sample of columns 4 and 5, where we exclude those who report having difficulty with the experimental instructions. In columns 3 and 6, worry about climate risks is not significantly correlated with wishful thinking. We note that our results on the marginally significant positive correlations are exploratory, subject to concerns about multiple hypothesis testing, and that they should be confirmed by future research.

7 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. 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 inference task and context in which beliefs are formed. Finally, we show that individuals differ in their propensity to engage in wishful thinking, with some showing the opposite tendency that reflects defensive pessimism.

Our results 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.[25] They help explain why people seek solace and comfort 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 false but reassuring 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. The subtle findings on the role of accuracy incentives indicate that the bias can persist despite high personal costs.

Our results point to number of open questions. For instance, it would be interesting to explicitly

[25]Examples are beliefs about health risks (Schwardmann, 2019), financial decisions and excessive trading (Brunnermeier and Parker, 2005; Eisenbach and Schmalz, 2015; Bridet and Schwardmann, 2017), time inconsistency (Caplin and Leahy, 2001; Köszegi, 2010), occupational choice and the labor market equilibrium (Akerlof and Dickens, 1982; Santos-Pinto et al., 2018), information acquisition (Yariv, 2002; Eliaz and Spiegler, 2006; Loewenstein, 2006), principal-agent communication (Köszegi, 2006; Caplin and Leahy, 2004), self-image and taboos (Bénabou and Tirole, 2011), groupthink (Bénabou, 2013) and politics (Bénabou, 2008; Levy, 2014; Le Yaouanq, 2016).

compare wishful thinking in the loss and gain domains, as loss aversion suggest that motives for wishful thinking will be stronger in the former domain. More generally, it would be important to understand the conditions under which wishful thinking responds to differences in losses or psychological stakes at the margin. It would also be interesting to look at whether the option to take a (costly) action to avert the adverse outcome, i.e. empowerment, can reduce wishful thinking. Finally, our results on how wishful thinking responds to material incentives suggest that cognitive investments in the accurate representation of signals are a key mechanism in motivated belief formation. Theoretical models, which have by and large focused on how material and psychological incentives affect ex-post belief distortions, may benefit from taking this mechanism seriously.

References

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

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

Auriol, Emmanuelle, Julie Lassebie, Amma Panin, Eva Raiber, and Paul Seabright, "God insures those who pay? Formal insurance and religious offerings in Ghana," *Mimeo, Toulouse School of Economics*, 2017.

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

Barron, Kai, "Belief updating: Does the 'good-news, bad-news' asymmetry extend to purely financial domains?," *WZB Discussion Paper*, 2016, (309).

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

Bénabou, R and J Tirole, "Self-confidence and personal motivation," *The Quarterly Journal of Economics*, 2002, (August), 871–915.

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

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

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

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

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

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

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

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

Bridet, Luc and Peter Schwardmann, "Selling Dreams: Endogenous Optimism in Lending Markets," *Mimeo, Ludwig Maximilians Universität München*, 2017.

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

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

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

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

Caplin, A and J Leahy, "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics*, 2001, (February 2001), 55–79.

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

— **and** — , “Wishful Thinking,” *NBER working paper 25707*, 2019.

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

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

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

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

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

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

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

Eisenbach, Thomas M. and Martin C. Schmalz, “Amxiety, overconfidence and excessive risk taking,” *Staff Report no. 711, Federal Reserve Bank of New York*, 2015.

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

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

Engelmann, Jan B., Friederike Meyer, Christian C. Ruff, and Ernst Fehr, “The neural circuitry of affect-induced distortions of trust,” *Science Advances*, 2019, *5* (3), eaau3413.

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

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

Falk, Armin and Florian Zimmermann, "Beliefs and utility: Experimental evidence on preferences for information," 2016.

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

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

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

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

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

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

Hollard, Guillaume, Sébastien Massoni, and Jean Christophe Vergnaud, "In search of good probability assessors: an experimental comparison of elicitation rules for confidence judgments," *Theory and Decision*, 2016, *80* (3), 363–387.

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

Islam, Marco, *Motivated Risk Assessments*, Department of Economics, School of Economics and Management, Lund University, 2021.

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

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

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

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

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

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

Le Yaouanq, Yves, "A model of ideological thinking," *Mimeo, Ludwig Maximilians Universität München*, 2016, pp. 1–49.

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

Leong, Yuan Chang, Brent L Hughes, Yiyu Wang, and Jamil Zaki, "Neurocomputational mechanisms underlying motivated seeing," *Nature Human Behaviour*, 2019.

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

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

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

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

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

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

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

Mayraz, Guy, "Wishful Thinking," *CEP Discussion Paper*, 2011, (1092).

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

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat, "Managing Self-Confidence," *Mimeo, Stanford University*, 2014.

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

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

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

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

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

Orhun, A Yesim, Alain Cohn, and Collin Raymond, "Motivated Optimism and Workplace Risk," *Available at SSRN*, 2021.

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

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

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

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

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

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

Schwardmann, Peter, "Motivated health risk denial and the market for preventative health care," *Journal of Health Economics*, 2019, *In press.*

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

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

Sharot, Tali, Alison M Riccardi, Candace M Raio, and Elizabeth A Phelps, "Neural mechanisms mediating optimism bias," *Nature*, 2007, *450* (7166), 102–105.

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

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

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

Sinding Bentzen, Jeanet, "Acts of God? Religiosity and Natural Disasters Across Subnational World Districts," *Economic Journal*, 2019, *In press.*