Crying Wolf in the Lab

Arya Gaduh, Peter McGee, Alexander Ugarov*

January 9, 2023

Abstract

Abstract is here —-

Keywords: alarms, value of information, information economics, information design, —

1 Introduction

The 2010 gas blowout on Deep Horizon oil rig has killed 11 workers and caused one of the largest oil spills in history. The death toll was possibly aggravated by switching off a general safety alarm because its sirens interfered with workers' sleep.[1] This illustrates the trade-off between false-positive and false-negative test results with false-positive rates leading to higher false alarm costs and false-negative resulting in missed events.

Many real-life situations involve choosing binary tests to discover and prevent a negative outcome. Most binary tests transform continuous signals about the likelihood of an adverse state into simple yes/no prediction. This transformation relies on choosing a threshold for positive classification. Holding a continuous signal constant, a decrease in probability of no alarm in an adverse state (false-negative rate) corresponds to an increase in probability of alarm in a non-adverse state (false-positive rate). This trade-off motivates multiple discussions in medical diagnostics, alarm systems and extreme weather alerts. Despite ubiquity of binary alarms, there is little empirical evidence on how users evaluate alarms with different false-positive and false-negative rates.

In order to understand preferences over these trade-offs, we study the demand for information in the framework with a potential protection action. The subject, first, receives a signal about the probability of an adverse event. Then she decides to protect or not. This environment describes several practically important scenarios including extreme weather alerts, medical testing and safety alarms.

Some recent studies observe that many people put non-zero value on information about ego-relevant beliefs or future utility even if it has no apparent effect on subsequent decisions (all the citations). These preferences is not the focus of our study and hence we use relatively low stakes and ego-neutral information. As a result, our findings might not apply to settings with changing identity beliefs or to settings with delayed resolution of uncertainty and large potential payoffs.

We find that the value of information in our setup weakly correlates with the willingness-to-pay. First, subjects on average underreact to quality of the signal, resulting in overpaying for low-quality signal and underpaying for high-quaity signals. Second, subjects tend to overreact to false-negative rates when the prior probability is low and overreact to false-positive positive rates when priors are high. We show that this pattern is most consistent with failure to estimate the effect of frequencies of false-positive and false-negative outcomes on the costs of using the signal. Xu (2020) similarly finds that individuals(?) do not properly account for priors and often choose tests not affecting optimal decisions even then more instrumental tests are available.

Our work is one of a few experimental studies measuring demand for information used for decision-making (instrumental information). Previous experimental studies studies the demand for signals in the prediction game in which subjects have to choose an optimal state under

[1] https://www.nytimes.com/2010/07/24/us/24hearings.html

uncertainty. The field experiment conducted by (Hoffman, 2016) finds that the demand for information increases with initial uncertainty, but decreases with the signal's accuracy. However, the decrease in accuracy is more modest than expected for a Bayesian decision-maker resulting in subjects underpaying for high-quality signals. The laboratory experiment of Ambuehl and Li (2018) finds that subjects tend to underreact to the accuracy of the binary signal about state of the world, but put a premium on completely certain signals. The paper of Xu (2020) similarly employs a prediction game but varies priors on top of signal charasteristics. reducing prior uncertainty makes more signals non-instrumental in the sense that there should be no effect from a signal to optimal decisions. She find that many subjects choose non-instrumental over instrumental signals which is consistent with

Our setup differs in two important aspects from (Ambuehl and Li, 2018; Xu, 2020), because we study alerts and not prediction tasks. The subject faces a costly protection decision and not a prediction decision, resulting in three distinct payoffs: full payoff, full payoff minus protection costs and full payoff minus losses. It means that risk preferences affect the value of information and can change sensitivities to false-positive and false-negative rates. Our findings however are similar to prediction game findings. Consistent with Ambuehl and Li (2018) we also find that subjects undervalue accurate signals, but we do not find a premium for certain signals. And similar to Xu (2020) we find that subjects do not properly account for interaction between prior probabilities and signal characteristics.

Due to its applicability for studying preferences over expectations, there is a larger stream of literature on the demand for non-instrumental information. Eliaz and Schotter (2010) find that subjects are willing to pay for signals even when these signals are excessive for making optimal choices. Their design involves subjects choosing between two boxes with one box containing a prize of \$20. Most subjects pay just to know the probability of finding \$20 in box A even if this box is more likely to contain a prize in all the possible states. This finding is inconsistent with expected utility maximization but indicates instead having preferences for certainty before making choices. Similar to this paper, Masatlioglu et al. (2017) also study preferences over information structures differing which differ in false-positive and false-negative rates but in their setup allows for a larger role of expectations. They find that for a positive potential outcome, most subjects prefer facing high false-negative rates rather than high false-positive rates. In other words, they tolerate uncertainty after negative signals better than uncertainty after positive signals. These preferences are salient: subjects require an average payment of 18-35 cents to switch to their least preferred information structure.

There is some mixed evidence that people update beliefs differently when these beliefs are ego-relevant or concern future gains and losses. Eil and Rao (2011) find asymmetry in updating ego-relevant beliefs such as beauty and IQ. Subjects update more after receiving positive signals and do not update enough after negative signals. Additionally, subjects with high posterior ego-relevant beliefs are willing to pay to receive a more precise signals, but require a compensation for learning when their beliefs are low. In contrast, Coutts (2019) does not find any updating

asymmetry with respect to either ego-relevant beliefs or beliefs about future payoffs.

Our paper is the first to measure value of information in the experimental setting of diagnostic tests or alarms. Previous work studies the use of alarms in context of medical testing, medical monitoring, safety alarms and extreme weather. Early literature on decision-making of medical professionals finds that doctors suffer from multiple biases when ordering testing, including inaccurate posterior probability estimation due to availability heuristics, hindsight bias and regret (Bornstein and Emler, 2001). Gigerenzer et al. (2007)find that very few mammologists understand mamogram results and tend to overestimate probability of cancer based on a positive result. Providing practitioners with natural frequencies instead of probabilities tends to reduce this bias.

Patients' willingness-to-pay for medical tests is large and largely responsive to test accuracy (Liang et al., 2003; Howard and Salkeld, 2009; Neumann et al., 2012). But there are several apparent violations of rationality. First, users are willing to pay for tests having little or zero diagnostic value (Schwartz et al., 2004; Neumann et al., 2012). For example, Schwartz et al. (2004) find that 73% of Americans in their survey prefer a free full-body CT scan versus one thousand USD cash. However, medical professional do not recommend full-body CT scans for healthy people due to extreme likelihood of false-positive findings. Second, the framing of test accuracy seems to matter a lot. Howard and Salkeld (2009) conduct a discrete-choice experiment to measure willingness-to-pay for the colorectal cancer screening. Their subjects agree to get 23 unnecessary colonoscopies in order to find one additional true cancer, but only 10.4 for reducing the number of cancers missed by one even though these descriptions are equivalent. Surprisingly, the perceived risk of cancer (prior) did not significantly affect the WTP in their study though the effect may come from its relatively low variation in the population.

This work also relates to the vast literature on demand for insurance and protection. Similar to our findings, several studies observe that the demand for insurance goes up after the recent experience with low-probabiity events. Field evidence indicates that people underinsure with respect to rare natural disasters (Friedl et al, 2014). Laury et al. (2009) find no under-insurance for low-probability events in the laboratory setting. One offered explanation (Volkman-Wise, 2015) is that subjects overweight recent evidence leading to underinsurance when there were no negative events in the recent past and to overinsurance after the fact. It is consistent with underweighting prior probabilities relative to more recent signals.

The bias we are finding is similar to the base-rate and signal neglect phenomenons. Psychology researchers Hammerton (1973) and Kahneman and Tversky (1973) first observed that subjects underweighted prior probabilities (base rates) when calculating posteriors. This phenomenon had received the name of *base-rate neglect*. Multiple studies in economics then confirmed (Grether, 1992; Holt and Smith, 2009) this phenomenon in incentivized laboratory experiments. Most of these studies find that subjects also underweight signals on top of priors. We observe both phenomenons in responses to our belief elicitation task, but the calculation

of signals' values differs substantially from the calculation of posterior probabilities. While the calculation of posterior probabilities would require using a Bayes formula, signal's value depends only on products of prior probabilities. However, we observe that subjects underestimate the effect of priors compared to theoretical predictions for an expected-utility decision-maker.

2 Model

Environment. The model describes a decision-maker considering a purchase of threat-assessment information. Let $\omega \in \{0, 1\}$ denote the state of world, where 1 corresponds to some adverse event happening with probability π. The decision-maker has a lower utility in the adverse state, but only if she does not take the protective action. Denote actions by $a \in \{0, 1\}$ with 1 meaning taking the protective action. The protection technology is perfect: protected agents bear no losses but pay protection costs c regardless of the state ω. Decision-maker preferences are described by the utility function which depends on wealth Y, protective action a and potential damage in the adverse state $\omega(1 - a)$. Utility is separable in wealth, protection costs $c > 0$ and potential loss in the adverse state $L > c$ [2]:

$$U = U(Y, a, \omega(1 - a)) = u(Y - ac - \omega(1 - a)L) \tag{1}$$

The decision-maker can purchase a binary informative signal $s \in \{0, 1\}$ about the state of the world before making a decision. Let $P_{ij} \equiv P(s = i|\omega = j)$ be the probability of a signal taking value i conditional on the state of the world being j. After receiving the signal, the decision-maker updates her belief on the likelihood of the bad state to $\mu(s)$. Unless specified otherwise, we assume that the decision-maker forms her posterior beliefs by using the Bayes rule. Hence the posterior belief equals:

$$\mu(s) = \frac{\pi P_{s1}}{\pi P_{s1} + (1 - \pi)P_{s0}} \tag{2}$$

We also assume without loss of generality that a higher signal means a higher posterior probability of an adverse event $\mu(1) \geq \mu(0)$. Otherwise we can always re-label the signals.

Preferences. If there is no signal, the decision-maker protects if and only if it increases their expected utility:

$$EU_0 = \max[u(Y - c), \pi u(Y - L) + (1 - \pi)u(Y)] \tag{3}$$

[2]Separability condition does not impose additional restrictions on the utility function U as long as the variation in wealth has limited range. More specifically, if $Y \in [Y_{min}, Y_{max}]$ and $c < Y_{max} - Y_{min}, L < c + (Y_{max} - Y_{min})$, then the function $u(\cdot)$ can be constructed from segments of $U(\cdot, 0, 0), U(\cdot, 1, 0), U(\cdot, 0, 1)$. While the resulting function $u(\cdot)$ is not necessarily monotonic, it is likely to be monotonic if protective actions and potential damages are relatively high.

The signal can increase expected utility if the decision-maker reacts differently to positive and negative signals. Under these assumptions, her expected utility with a signal is:

$$EU_s = \pi P_{11} u(Y-c) + \pi P_{01} u(Y-L) + (1-\pi) P_{10} u(Y-c) + (1-\pi) P_{00} u(Y) \tag{4}$$

We consider the maximum amount b which the decision-maker is willing to pay for the signal. In our framework, it is a price paid with a signal such that a decision-maker is indifferent between having a signal and paying b and not having a signal. Because the decision-maker can always ignore a useless signal, the signal's value is bounded from below by zero. Hence it equals to the maximum between zero and the solution to the following equation:

$$\begin{aligned} P(s=1)u(Y-b-c) + \pi P_{01} u(Y-b-L) + (1-\pi) P_{00} u(Y-b) = \\ = \max[u(Y-c), \pi u(Y-L) + (1-\pi) u(Y)] \end{aligned} \tag{5}$$

The left-hand side expression of this equation is a strictly decreasing function of b. Additionally, for $b \to \infty$ the left-hand side is smaller than the right-hand side. It implies that the equation (5) above has a at most one positive solution.

Obviously, perfectly accurate signals always have positive value $b > 0$ because the payoff distribution with the signal first-order stochastically dominates the distribution without the signal. If the decision-maker protects without a signal, a perfect signal reduces the protection costs and if she takes chances, then it reduces losses in the adverse outcome from L to $c < L$. However, it is harder to determine the value of the imperfect signal without imposing more restrictions on preferences as it requires weighing $u(Y-L)$ against $u(Y-c)$.

Risk-neutral agent. If the decision-maker is risk-neutral, the expression above collapses to:

$$b + P(s=1)c + \pi P_{01} L = \min[c, \pi L]$$

The signal's value is just:

$$b = \max[0, \min[c, \pi L] - P(s=1)c - \pi P_{01} L] \tag{6}$$

We can express WTP b as a function of priors, false-positive and false-negative rates. This is the equation we use in our empirical work:

$$b = \max[0, \min[c, \pi L] - \pi(1-P_{01})c - (1-\pi)P_{10}c - \pi P_{01} L] \tag{7}$$

The sensitivity of (positive) value b with respect to false-positive and false-negative rates is given by:

$$\frac{db}{dP_{10}} = -(1-\pi)c \tag{8}$$

$$\frac{db}{dP_{10}} = -\pi(L - c) \tag{9}$$

Both false-positive and false-negative rates decrease the (positive) signal's value. The effect is proportional to the adverse state probability for the false-negative rate and to the non-adverse state probability for the false-positive rates.

Risk Aversion Effects. In a more general expected utility framework, risk aversion can both increase and decrease the signal's value. More specifically, risk aversion decreases the value when the protection costs are low:

Proposition 1. *If protection costs are low $c < \pi L$, then the strictly risk-averse decision-maker pays less than a risk-neutral one.*

Proof. See the Appendix. □

It is harder to make definite statements for lower risks or higher protection costs. For example, risk aversion increases value of a perfect signal as long as risk-averse decision-maker still chooses to not protect without a signal. This follows from the standard argument of increasing demand for insurance with risk aversion and the fact that the protection problem with a perfect signal is isomorphic to the insurance problem with deductible c.

Next, we study the effect of false-positive and false-negative rates on the signal's value b. Assuming a differentiable utility function $u()$ we use implicit differentiation to derive sensitivities of WTP b to false-positive and false-negative rates:

$$\frac{db}{dP_{10}} = -\frac{(1-\pi)(u(Y-b) - u(Y-c-b))}{D(\pi, P_{01}, P_{10}, b)}$$

$$\frac{db}{dP_{01}} = -\frac{\pi(u(Y-c-b) - u(Y-L-b))}{D(\pi, P_{01}, P_{10}, b)}$$

With the denominator equal to the expected marginal utility:

$$D(\pi, P_{01}, P_{10}, b) \equiv P(S=1)u'(Y-c-b) + \pi P_{01}u'(Y-L-b) +$$

$$+(1-\pi)P_{00}u'(Y-b) = E[MU] > 0$$

It is clear that the signal's value decreases with false-positive and false-negative rates $\frac{db}{dP_{10}}, \frac{db}{dP_{01}} < 0$. We can also say a bit more about the sensitivity to false-negative rates:

Proposition 2. *Risk-averse and imprudent decision-maker has higher sensitivity to false-negative rates as compared to a risk-neutral one.*

Proof. Use the mean value theorem to rewrite the sensitivity as:

$$\frac{db}{dP_{01}} = -\frac{\pi u'(\zeta)(L-c)}{E[MU]}, \zeta \in (Y-c-b, Y-L-b)$$

Now let X denote a random payoff of the decision-maker with a signal. A risk-averse decision-maker puts a positive value on the signal only if its expected payoff is higher than the payoff with full protection: $EX > Y-c-b$. If a decision-maker is imprudent ($u''' < 0$) then $E[MU] \equiv E[u'(X)] < u'(EX)$. Next, because u' is a strictly increasing function and $EX > Y-c-b$: $u'(\zeta) > u'(Y-c-b) > u'(EX)$. Hence $\frac{u'(\zeta)}{E[MU]} > 1$ and $\frac{db}{dP_{01}} < -\pi(L-c)$. □

However, risk aversion can both increase and decrease subject's sensitivity to false-positive rates depending on the utility function curvature and signal's characteristics. Intuitively, an expected marginal utility of a strongly risk-averse subject with a bad signal can be lower than the average slope of the utility function between $Y-c-b$ and $Y-b$ reducing sensitivity to false-positive rates. It can also be higher if either the signal is good or the curvature is small.

3 Experimental Design

In each session, subjects received a USD 5 show-up fee and were endowed with USD 25 that they might lose in the experiment. Subjects must then make a series of decisions in four sets of tasks: (i) Blind Protection; (ii) Informed Protection; (iii) Belief Elicitation; and (iv) Willingness to Pay Elicitation. To verify that subjects understand these tasks, they must answer a quiz before each task. If a subject gets any answer wrong, they read correct answers and explanations for each wrong answer. Additionally, subjects receive extra questions if they give wrong answers in a 5-question quiz given before the Informed Protection task. We do this because we consider Informed Protection as a first challenging task in the sequence which understanding is essential for the rest of the tasks. Each set of tasks has 6 rounds, for a total of 24 rounds. One of these rounds is selected at random as the payment round. A copy of the instruction is included in Appendix XX.

Blind Protection (BP). In each BP round, subjects must decide whether to insure (or "protect") against an adverse event (i.e., drawing a black ball from a box). Subjects were informed of the prior probability of drawing a black ball before making their decision. The cost to protect is USD 5. If a black ball is drawn, an unprotected subject will lose USD 20. Subjects then played six rounds, where the probability of drawing a black ball was varied between XX and XX percent in each round. During the BP task, subjects did not receive any feedback on how that round would have been realized were it chosen as the payment round.

Informed Protection (IP). For the IP task, subjects make a protection decision as in BP. However, before each decision, subjects are given a signal that was generated with varying

degrees of inaccuracy. Following Coutts (2019), we present the signal-generation process using groups of "gremlins" that represent three types of signals: accurate (an honest gremlin), false positive (a black-swamp gremlin that always announces that the ball is black), and false negative (a white-swamp gremlin that always announces that the ball is white). Figure XX illustrates how the different gremlin types were presented to the subjects. Subjects knew the composition of the group from which the hint came from, but did not know which gremlin provided the hint. We vary the proportion of black balls in the box (prior probability of a black ball) and the composition of gremlins (signal quality) between rounds.

Belief Elicitation (BE). We use the BE task to elicit subjects' beliefs about the likelihood of an adverse event and an adverse signal conditional on prior and signal characteristics in an incentive-compatible way. Similar to the IP task, subjects were informed of the prior probability of a black ball and the composition of the group of gremlins that would provide an additional signal. However, instead of asking subjects to make a protection decision, we asked them to estimate the probability of two events, to wit: (i) the ball is black ball when a randomly drawn gremlin says that it is white; (ii) the ball is black when a randomly drawn gremlin says that it is black.

We follow the stochastic version of the Becker-DeGroot-Marshak mechanism developed by Grether (1992) and Holt and Smith (2009) to elicit incentive-compatible responses: the subject submits their belief of the probability of the event $\mu \in [0, 1]$. If this belief is above some uniform random number $r \in [0, 1]$, they receive the payoff x only if the stated event happens. Otherwise their payoff is determined by an independent lottery which pays x with probability r and 0 otherwise.[3] To help subjects understand this complex mechanism, we prefaced our explanation of it with the fact that under this mechanism, truthful reporting of beliefs is the dominant strategy.

Willingness to Pay Elicitation (WTPE). The WTPE task measures subjects' willingness to pay (WTP) for signals. Subjects know the prior probability of a black ball and the group composition of the gremlins that will determine signal quality. We then ask subjects for their WTP to receive a hint from a randomly drawn gremlin. Subjects can choose a value from USD 0 to 5 with USD 50-cent increments. Their decisions are incentive compatible: if a WTPE round is selected as the payment round, a random price of a hint will be drawn. If that price exceeded the subjects' WTP, they will play a BP round. Otherwise, the subject would pay for the hint and play an IP round. After completing the WTPE task, subjects were asked a few demographic questions. The session concluded with the random selection and realization of the payment round, after which subjects were paid and dismissed.

[3]The benefit of this mechanism versus other probability elicitation mechanism (for example, quadratic scoring) is that reporting truthfully is a dominant strategy regardless of risk preferences (Karni, 2009). The only requirements a subject needs to satisfy are probabilistic sophistication and dominance: they rank lotteries based on their probabilities only and prefer higher probabilities of higher payoffs.

The first three tasks were designed to provide measures of the different components of WTP described in Section XX and use them to examine the extent to which they explain subjects' WTP measured in the WTPE task. We use the BP task both to measure subjects' responses to the prior and their risk aversion. Next, we use the IP task to examine how signals affect protection decisions. Finally, we use the BE task as a measure of subjects' ability to estimate the probability of a signal for a given quality and to perform Bayesian updating. To construct these measures, we presented our subjects with 6 different priors for the BP task, and 3 priors and 2 gremlin groupings for the IP, BE, and WTPE tasks. Table reftab:treatments XX shows the values of the different priors in our treatments, as well as the gremlin groupings (along with the associated false positive and false positive rates) that we used for the different tasks.

We conducted this experiment in the Behavioral Business Research Lab (BBRL) at the University of Arkansas between October and November 2021. The experiment was implemented using Qualtrics. There were a total of 105 subjects. 84 percent of the subjects were university students and 41 percent were male. About 60 percent of the subjects had taken at least one statistics course. On average, including the show-up fee, subjects received around USD 26 for a session lasting around 45 minutes.

Table 1: List of Treatments

	Gremlins composition				
Prop. of black balls (p)	Honest	Black-eyed	White-eyed	FP rate	FN rate
0.1, 0.2, 0.3, 0.5	2	0	0	0	0
0.1, 0.2, 0.3, 0.5	3	1	0	0.33	0
0.1, 0.2, 0.3, 0.5	3	0	1	0	0.33
0.1, 0.2, 0.3, 0.5	3	1	1	0.33	0.33
0.1, 0.2, 0.3, 0.5	5	1	0	0.2	0
0.1, 0.2, 0.3, 0.5	5	0	1	0	0.2
0.1, 0.2, 0.3, 0.5	5	1	1	0.2	0.2

4 Results

We begin with a brief overview of subject behavior in the different tasks.

XXXX NEED TO REWRITE HERE XXXX

We follow that discussion with a regression analysis to explain subjects' WTP for signals of different qualities. Our regression results suggest that subjects' WTP deviated from those of a risk-neutral utility maximizing subject which was driven by their failure to fully account for signal quality when calculating their WTP. Furthermore, we find that these deviations remained after controlling for risk aversion or subjects' ability to perform Bayesian updating.

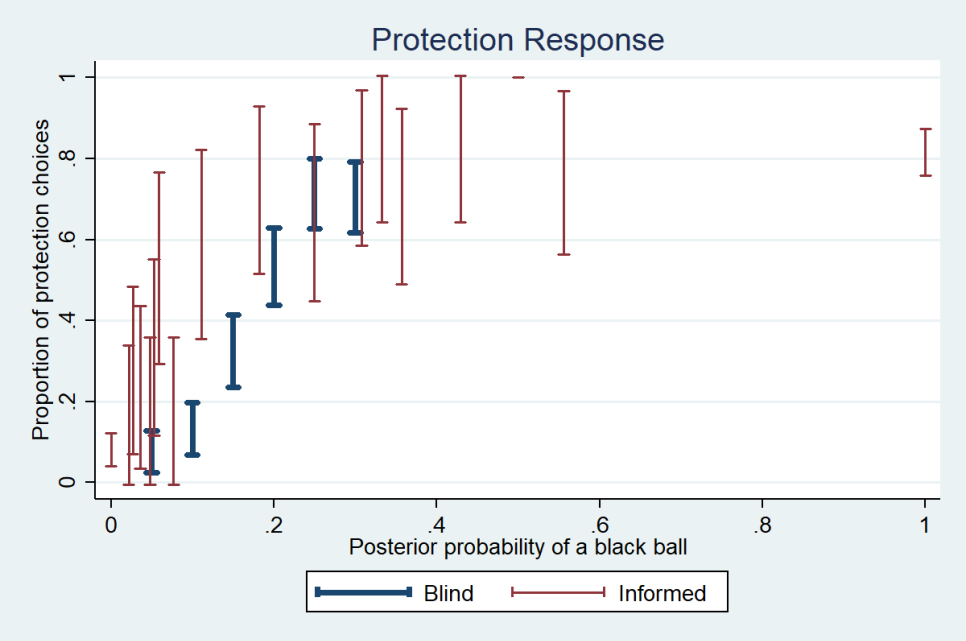
4.1 Overview: Are Subjects Bounded Rational?

Risk Aversion. Subjects' responses in the BP task are generally consistent with the expected utility framework. Figure 1 shows the probability of protection decision given prior probability of a black ball from the BP task. Subjects protected more with a higher probability of a negative outcome: only 13% subjects protect when the probability of a black ball is 10% in contrast to 70% protecting when the probability is 30%. About 30% of subjects vacillated between protecting and not protecting; however, most of them (24 out of 39) vacillated only once, skipping a single increment of the presented probability scale, suggesting an inattention error. Following XX ALEX: Any cite? XX, we "corrected" these single-skip BP responses when calculating subjects' risk aversion.

BP responses indicate significant heterogeneity in terms of risk aversion. Risk-neutral subjects maximized their expected utility by protecting whenever the prior probability exceeds 0.25, which is the ratio of the protection cost (USD 5) to the potential loss (USD 20). In contrast, many subjects started protecting for lower probabilities of 0.1 or 0.2 indicating risk aversion. A smaller group of subjects makes choices consistent with risk loving by never protecting or protecting for the probability of 0.3.

Protection Response to Signal. We use the IP task to examine how subjects responded to a signal for given priors and signal quality. Since signal quality is known, we can calculate the true posterior probability of drawing a black ball. Figure 1 presents the subjects' responses in the IP task plotted against these true posteriors. Consistent with their behavior in the BP task, the share of subjects who protected in the IP task is increasing in the posterior probabilities. Subjects' responses in the first two tasks suggest that conditional on the true probabilities, subjects protected more in the IP task compared to the BP task at low probabilities. It appears that subjects' updated beliefs overshot the true posterior at low probabilities. It is also consistent with ambiguity aversion in the IP task. However, we cannot use these tasks to make general inferences about subjects' ability to update since the BP task did not cover the full range of probabilities.

Figure 1: Average Protection Response



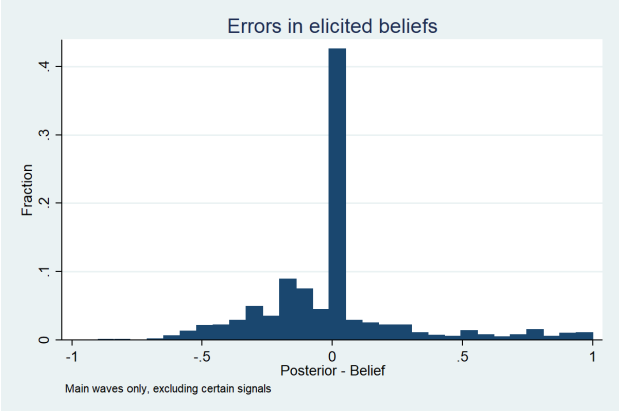
Bayesian Updating Accuracy. Since subjects were only given signal characteristics and not true posteriors, their IP responses reflect, inter alia, their ability to infer the true posteriors from signal characteristics. Figure 2 presents subjects' abilities to update their beliefs based on the BE task. We define updating errors as the difference between the posterior and subjects' elicited belief on the posterior probability of a black ball for a given signal. Figure 2a presents the distribution of these differences. The distribution of updating errors is centered at 0. Figure 2b presents the scatter plot of the elicited belief and the true posterior probability for a given signal. The correlation between elicited belief and the true posterior was XXXX .

We can group the treatments in the BE task into: (i) those where the posterior is uncertain (between zero and one); and (ii) those where the posterior is certain (either zero or one). In Figure 2c, we plot the distribution of updating errors for the first group of treatments. Its mean is negative, suggesting that subjects tend to overestimate the likelihood of adverse events for uncertain posteriors. The correlation between their belief and the true posterior in this subset of observations is XXX.

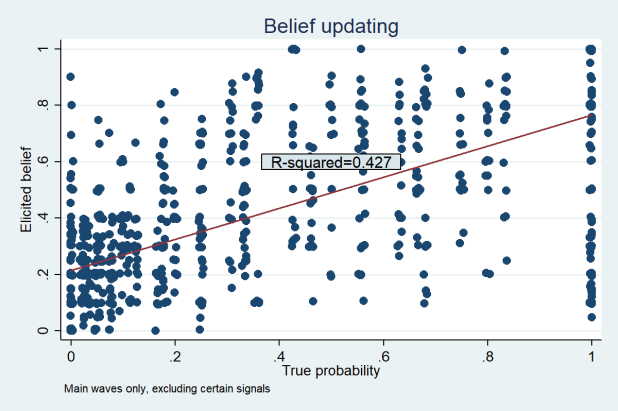
Figure 2e plots the distribution of updating errors with certain posteriors, which includes: (i) treatments with all-honest gremlins; and (ii) treatments with dishonest gremlins (e.g., a group with honest and white-eyed gremlins with a hint that the ball is black — or vice versa). With perfectly honest signals, responses were incorrect for about 20 percent of the time. About half of these involve reporting a probability of between one and zero, with the other half reporting a probability of one when it should have been zero. For the latter case, only 51 percent of responses were correct.

Figure 2: Errors in Bayesian Updating

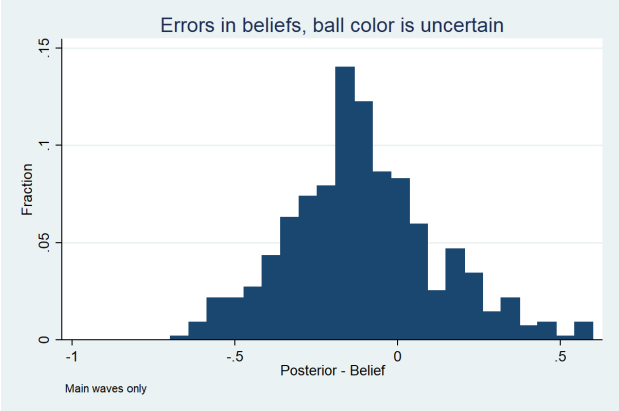
(a) Error Distribution



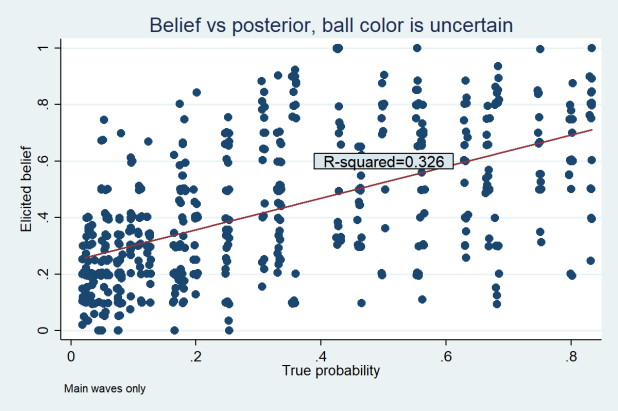
(b) Error v. Posterior



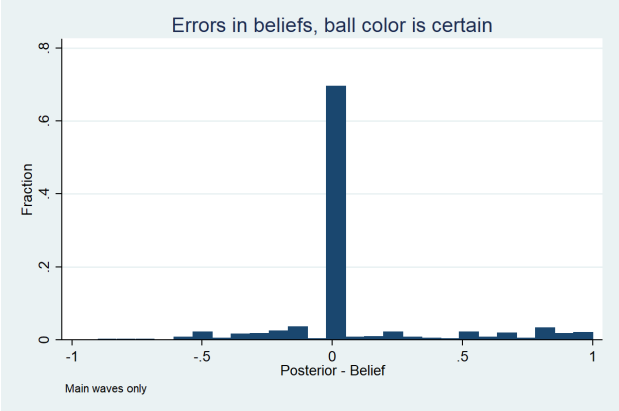
(c) Error Distribution, Uncertain Color



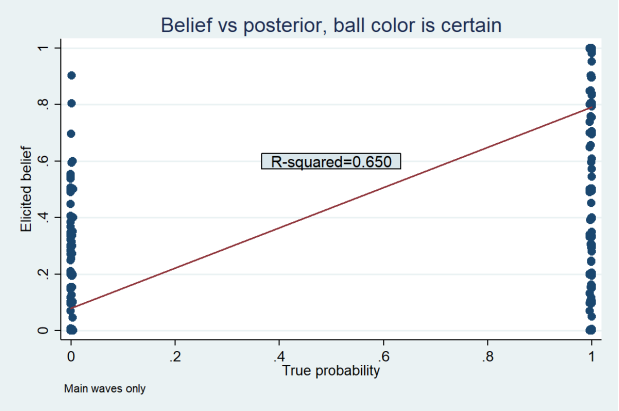
(d) Error v. Posterior, Uncertain Color



(e) Error Distribution, Certain Color



(f) Error v. Posterior, Certain Color



4.2 Signal Characteristics and Protection Decision

XXX WE NEED TO JUSTIFY SOMEWHERE THAT WE LIMIT OUR SAMPLE TO 0.1, 0.2 FOR THE MAIN ANALYSIS STARTING FROM HERE XXX

Hypothesis 1. *Conditional on posterior XXX and risk preferences?? XXX, signal characteristics do not affect protection decisions.*

Result 1. *Signal characteristics affect protection decision. Conditional on posterior XXX and risk preferences?? XXX, subjects' protection decisions still respond to the signals' false positive and false negative rates.*

Table 2 presents a non-parametric analysis of subjects' IP decisions by signal characteristics (columns 1–3). Column 4 shows the XXX ALEX: How do we construct the posterior for multiple treatments? Averaging them? XXX . Column 5 shows the share of IP responses that chose to protect, with column 6 showing the p-value of a t-test that nobody protected. Column 7 shows... XXX ALEX: Also how do we calculate the optimal? Please add explanation XXX, with column 8 showing the p-value of a t-test that the share who protected is equal to the optimal share.

Overall, subjects' protection decisions deviated from what are optimal for the risk-neutral subject and these deviations systematically depend on FP and FN rates. These deviations systematically different between white and black signals. Subjects overprotected (compared to the optimal) in response to FP rates when the signal is white (rows 1–4). For example, consider the change in the protection rates between rows 1 and 3: because the signal is white, an increase in the signal's FP rate did not change the posterior. However, the protection rate increased by more than 11 percentage point (pp.). Similarly, row 4 shows that when both FP and FN are positive, the protection rate increased to 43 pp. — even though the average (maximum) posterior probability for the signal characteristics is 6 (11) percent. As a benchmark, only 13 (32) percent of subjects chose to protect in the BP task when the probability is 10 (15) percent.

Meanwhile, rows 5–8 show that when the signal is black, subjects protected less than what was optimal for a risk-neutral agent. However, unlike for the white signal, it appears that subjects are less responsive to changes in FP and FN rates.

Table 2: Average Protection by Signal Type

Row	Signal Characteristics			Posterior	Share Protect	P-val (H_0 : $ShProt = 0$)	Share Optimal	P-val (H_0 : $ShProt = ShOptimal$)
	False Positive	False Negative	Signal					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	No	No	White	0.000	0.038	0.022	0.000	0.045
2	No	Yes	White	0.045	0.188	0.000	0.000	0.000
3	Yes	No	White	0.000	0.145	0.001	0.000	0.001
4	Yes	Yes	White	0.062	0.429	0.000	0.000	0.000
5	No	No	Black	1.000	0.837	0.000	1.000	0.000
6	No	Yes	Black	1.000	0.783	0.000	1.000	0.000
7	Yes	No	Black	0.396	0.739	0.000	0.739	1.000
8	Yes	Yes	Black	0.328	0.829	0.000	0.743	0.182

Notes:

Hypothesis 2. *Subjects' Bayesian-updating errors explain IP decisions.*

Result 2. *When subjects received a signal that the ball is white, the signal's false positive and false negative rates biased their belief upward. When subjects received a signal that the ball is black, the signal's false positive (negative) rates biased their belief upward (downward). Updating errors provide partial explain for subjects' IP decisions conditional on posterior.*

Deviations from the optimal IP response can arise from subjects' failure to correctly update their posteriors. Table 2 summarizes how the updating errors vary with signal characteristics. We find that when presented with a white signal, subjects failed to fully downward-adjust the likelihood of a black ball, biasing their estimates of the posterior upward. Introducing FP rates to the signal exacerbated their upward bias. To illustrate, consider the change between rows 1 and 3, where introducing an FP rate would not change the posterior (= 0) since the signal is white. Yet, subjects updated their posterior upward, magnifying their updating error. The FN rates also have a similar effect of exacerbating this upward bias for a white signal.

Meanwhile, subjects underestimated the increase in posterior when the signal is black. As in the case of the white signal, introducing FP rates led subjects to overestimate the posterior instead. In contrast, introducing FN rates led subjects to further underestimate the posterior. To illustrate, we can examine rows 5 and 6, where the introduction of an FN rate given a black signal would not change the posterior — yet, subjects updated their posterior downward.

Table 4 formalizes our analysis using a regression. We estimated XXX ALEX: presumably these are linear but just to be sure, can you fill this in. Thanks! XXXX. It confirms our findings that: (i) subjects made positive (negative) updating errors for white (black) signals; (ii) FP rates induced an upward bias in subjects' estimates of the posterior; and (iii) FN rates induced an upward (downward) bias when the signal is white (black).

Table 3: Average Updating Error by Signal Type

Row	Signal Characteristics			Posterior	Updating Error*	P-val (H_0 : $Error = 0$)
	False Positive	False Negative	Signal			
	(1)	(2)	(3)	(4)	(5)	
1	No	No	White	0.000	0.039	0.001
2	No	Yes	White	0.045	0.140	0.000
3	Yes	No	White	0.000	0.116	0.000
4	Yes	Yes	White	0.062	0.245	0.000
5	No	No	Black	1.000	-0.187	0.000
6	No	Yes	Black	1.000	-0.332	0.000
7	Yes	No	Black	0.396	0.177	0.000
8	Yes	Yes	Black	0.328	0.192	0.000

Notes: *Updating error = $Posterior - Belief$.

Table 4: Updating Errors in BE Task

	All	Signal Received	
		White	Black
	(1)	(2)	(3)
FP rate	.948***	.318***	1.58***
	(0.1)	(0.1)	(0.1)
FN rate	.00702	.38***	-.366***
	(0.1)	(0.1)	(0.1)
Constant	-.249***	.139***	-.636***
	(0.0)	(0.0)	(0.0)
Observations	624	312	312
Adjusted R^2	0.22	0.37	0.66
Subject FE	Yes	Yes	Yes

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Certain aspects in the way subjects made updating errors are consistent with how they made IP decisions. In Table 5, we regress the following XXX ALEX, can you describe the model and specify? XXX:

$$MODEL$$

where ..... Columns 1 and 2 included flexible controls of the true posteriors.[4]. Columns 3 and 4 added further flexible controls to account for subjects' (often incorrect) estimates of the posterior, inferred from their BE responses.[5] The model is estimated using XXX, with standard errors clustered at the subject level. The table presents the average marginal effect coefficients.

[4]XXX ALEX: Describe what the flexible controls are in the footnote XXX

[5]XXX ALEX: Describe how the controls for beliefs are constructed, also the flexible aspect. XXX

Columns 1 and 2 confirmed Result 1, to wit, conditional on posterior and risk preferences?? subject FE??? XXX, IP responses are affected by FP and FN rates. For a white signal, FP and FN rates increased the tendency to overprotect; while for a black signal, only the FP rate had a similar effect XXX ALL: I am just eyeballing here, would need to formally test and just discuss the result of the test in the footnote XXX. Column 3 suggests, however, that once we control for both the posterior and subjects' updated belief, only the effect of the FP rate for white signals persist. These results provide evidence that subjects' failure to protect optimally is largely — albeit not entirely — driven by their failure to correctly update their posterior given a signal.

XXX DO WE NEED COLUMNS 2 and 4? XXX

Table 5: Informed Protection Response

	(1)	(2)	(3)	(4)
FP rate	.523***	.488**	.369*	.282
	(4.0)	(2.0)	(1.9)	(1.1)
FN rate	.724***	1.36***	.512	.833**
	(4.6)	(3.4)	(1.3)	(2.0)
S=Black	.321**	2.4***	.731	1.8**
	(2.5)	(3.4)	(1.3)	(2.6)
FP rate x (S=Black)	-.119	-3.42***	-1.08	-2.5**
	(-0.4)	(-2.9)	(-1.1)	(-2.2)
FN rate x (S=Black)	-.721***	-1.64***	-.557	-1.14***
	(-3.6)	(-4.0)	(-1.4)	(-2.7)
p≥0.2	.119***	.351***	.35***	.299***
	(4.3)	(7.1)	(6.8)	(5.1)
FP rate x (p ≥ 0.2)		.573*		.409
		(1.7)		(1.2)
FN rate x (p ≥ 0.2)		.556**		.589**
		(2.3)		(2.1)
Observations	1224	582	582	582
Adjusted R^2				
Subject FE	Yes	Yes	Yes	Yes
Flexible controls for:				
Posterior	Yes	Yes	Yes	Yes
Beliefs	No	No	Yes	Yes

Notes: Coefficients are average marginal effects. t-statistics in parentheses. Standard errors are clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. XXX ALEX: WHY ARE THERE 1224 OBSERVATIONS IN COLUMN 1. ARE THESE ALL OBSERVATIONS, WHILE 2-4 ARE ONLY P=0.1/0.2? XXX

XXX ARYA STOPPED HERE XXX

4.3 Willingness to Pay

Figure XX shows how subjects deviate substantially from the theoretical value for a risk-neutral subject. Here we explore these deviations more systematically by signal characteristics and risk preferences.

Table 6: WTP for Information (tobit)

	(1) All	(2) p=0.1	(3) p=0.2	(4) All	(5) All	(6) All
model						
FN costs	-.577**	-1.24**	-.682***	-.791***	-.691***	-.69***
	(0.2)	(0.5)	(0.3)	(0.2)	(0.2)	(0.3)
FP costs	-.644***	-.647***	-.519**	-.595***	-.508***	-.494**
	(0.2)	(0.2)	(0.3)	(0.2)	(0.2)	(0.2)
BP costs				.373***	.363***	.37***
				(0.1)	(0.1)	(0.1)
Belief change					.332	
					(0.3)	
Certainty						.688
						(0.8)
Constant	1.98***	1.79***	2.33***	.923***	.701*	.293
	(0.2)	(0.2)	(0.2)	(0.3)	(0.4)	(0.8)
sigma						
Constant	1.8***	1.83***	1.7***	1.77***	1.76***	1.76***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	312	159	153	312	312	312
Adjusted R^2						

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table XX, we use a regression analysis to investigate how and why subjects deviate from the risk-neutral subjects' theoretical WTP for a signal of a given quality (conditional on prior). We first estimate how individual deviations from the theoretical benchmark b_s^* for a given signal s are correlated with signal's charactaristics:

$$b_{is} - b_s^* = \beta_0 + \beta_1 FalsePositive + \beta_2 FalseNegative + \epsilon_{is}$$

where b_{is} is the reported WTP of individual i for treatment s and b_s^* is the signal's value for a risk-neutral subject, and FalsePositive (FalseNegative) is the false positive (false positive) costs variables that captures signal quality. We calculate false positive costs as the product of prior probability of a white ball multiplied by the conditional probability of getting a black signal ("The ball is black!") from a randomly chosen gremlin: $FalsePositive = (1 - \pi)P_{10}c$.

Similarly we calculate false negative costs as the probability of an adverse state multiplied by a conditional probability of getting a white signal conditional on the ball being black and multiplied by potential loss $FalseNegative = \pi P_{01} L$. Note that these costs already account for expected frequency of receiving different incorrect signals as consistent with their base rate. If our subjects were risk neutral expected utility maximizers, we expect β_1 and β_2 to be zero.

Table 7: Average WTP discrepancy (WTP-Value) by Signal Type

False-positive	**False-negative**	**Mean WTP discrepancy**	**P(= 0)**
No	No	-0.135	0.465
No	Yes	-0.209	0.152
Yes	No	0.465	0.005
Yes	Yes	0.437	0.001

Column 1 of Table XX confirms that subjects' WTP deviated from the theoretical benchmark. Subjects did not fully account for signal quality, resulting in overpaying for signals with either high false-positive and false-negative costs. Naturally, two potential sources of deviations from this theoretical benchmark based on a risk-neutral Bayesian updater are subjects' risk-preferences and their ability to perform Bayesian updating. To test for these mechanisms, we interacted the false positive (FP) and false negative (FN) variables with individual risk aversion, whether individuals have accurate belief (as measured by our BE task), and the different priors.

Column 2 shows the results of the regression where the signal quality variables were interacted with the subject's risk preference.[6] This premium doesn't seem to come from risk aversion, as the coefficient on the interaction of risk aversion with FP and FN rates is relatively small and insignificant. Belief accuracy measured in the belief elicitation task apparently explains away the excess sensitivity to the FP rate but this finding should be taken with caution because the coefficient is not statistically significant despite its large absolute magnitude.

These results suggest that, on average, subjects failed to fully account for signal quality, resulting in overpaying for signals with high false-positive and false-negative costs. FP/FN significantly impact the deviation from the theoretical value no matter what else is included.

[6]Our risk preference estimates come from blind protection choice: subjects switching from no protection to protection at exactly the cost-loss ratio $\pi = 0.2$ are considered risk-neutral, while switching at lower (higher) levels indicates risk aversion (risk-loving).

Table 8: WTP minus Value of Information (OLS)

	(1)	(2)	(3)	(4)	(5)
FP costs	.558***	.472***	.403	.506***	.437***
	(0.1)	(0.1)	(0.3)	(0.2)	(0.1)
FN costs	-.229*	.0337	-.495	.085	-.645***
	(0.1)	(0.1)	(0.5)	(0.1)	(0.2)
Risk-loving × FP costs			.12		
			(0.4)		
Risk-averse × FP costs			.102		
			(0.3)		
No risk av. measure × FP costs			-.142		
			(0.4)		
Risk-loving × FN costs			.744		
			(0.5)		
Risk-averse × FN costs			.549		
			(0.5)		
No risk av. measure × FN costs			.492		
			(0.5)		
Inaccurate beliefs				.0776	
				(0.2)	
Inaccurate beliefs × FP costs				.631	
				(0.8)	
Inaccurate beliefs × FN costs				-.00734	
				(0.3)	
plevel=200 × FP costs					.14
					(0.2)
plevel=200 × FN costs					.84***
					(0.2)
Constant	-.0921	-.141*	-.137	-.208	-.111
	(0.2)	(0.1)	(0.1)	(0.2)	(0.1)
Observations	312	312	312	312	312
Adjusted R^2	0.05	0.59	0.58	0.58	0.60

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Summary

Table 9: Comparing Findings across the Tasks

Design	Beliefs	IP	WTP
White, FN only	>	<>	<> *
Black, FN only	<	<>	<>
White, FP only	>	>	>
Black, FP only	>	<>	>
White, FN and FP	>>	>	>
Black, FN and FP	>	<>	>

*-WTP estimates do not depend on signals.

4.5 Subject Heterogeneity

Table 10: Latent Class Multinomial Choice Model Estimates (FP and FN rates by hint)

	lc_results									
	Model	Class	Alt	Hint	FN0	FN1	FP0	FP1	Class share	
r1	1	1	-2.86694	4.392251	4.834518	-.1919326	4.35168	-.8676941	1	5
r2	2	1	-2.91958	1.881626	7.980388	-.3599557	1.725487	6.632253	.2198715	5
r3	2	2	-2.91958	6.699559	3.838407	.4707898	5.285504	-8.229022	.7801285	5

Table 11: IP response by class

	(1) Honesty Seekers	(2) Cautious Bayesians
S=Black	.337***	.0245
	(3.4)	(0.4)
Prop. of lying gremlins	.664***	.277***
	(4.6)	(4.3)
Posterior prob.	-.198*	.788***
	(-1.7)	(4.9)
N	138	486
Pseudo R-squared	.183	.541
Log-likelihood	-67.2	-154

t statistics in parentheses

Errors are clustered by subject, average marginal treatment effects

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Belief Elicitation by Class

	(1) Simpletons	(2) Cautious Bayesians
Posterior prob.	.357***	.479***
	(0.1)	(0.1)
S=Black	.123	.224***
	(0.1)	(0.0)
Prop. of lying gremlins	.171	.184***
	(0.1)	(0.0)
Constant	.112***	.0898***
	(0.0)	(0.0)
Observations	138	486
Adjusted R^2	0.31	0.60

Standard errors in parentheses

Dep. variable: beliefs, errors clustered by subject

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Expected IP losses by strategy

	p=0.1,0.2			p>0.2		
	Mean loss	% of optimal	Loss prob.	Mean loss	% of optimal	Loss prob.
Baseline (all)	1.166304	156.7689	.0190281	2.11717	140.6088	.0508233
Honesty seekers	1.526998	205.2517	.0435806	3.095308	205.5705	.1163925
Bayesians	1.050706	141.2308	.0112388	1.806053	119.9464	.0300237
Optimal	.7439637	1	.0136432	1.505716	1	.0190598

Table 14: Belief Elicitation: When Mistakes Happen

	(1)	(2)	(3)
	All	S=White	S=Black
Simpletons	.28***	-.105***	.665***
	(0.0)	(0.0)	(0.0)
FN rate	.0528	.409***	-.304**
	(0.1)	(0.1)	(0.1)
Simpletons × FN rate	-.177	-.0993	-.255
	(0.2)	(0.2)	(0.3)
FP rate	.888***	.253***	1.52***
	(0.1)	(0.1)	(0.1)
Simpletons × FP rate	.277	.316	.238
	(0.2)	(0.3)	(0.4)
Constant	-.251***	.14***	-.641***
	(0.0)	(0.0)	(0.0)
Subject FE	Yes	Yes	Yes
Observations	624	312	312
Adjusted R^2	0.22	0.38	0.66

Standard errors in parentheses

Dep. variable: reported belief - posterior probability

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5 Conclusion

References

Ambuehl, Sandro and Shengwu Li (2018) "Belief updating and the demand for information," *Games and Economic Behavior*, 109, 21–39, 10.1016/j.geb.2017.11.009.

Bornstein, B. H. and A. C. Emler (2001) "Rationality in medical decision making: a review of the literature on doctors' decision-making biases," *Journal of Evaluation in Clinical Practice*, 7 (2), 97–107, 10.1046/j.1365-2753.2001.00284.x.

Coutts, Alexander (2019) "Good news and bad news are still news: experimental evidence on belief updating," *Experimental Economics*, 22 (2), 369–395, 10.1007/s10683-018-9572-5.

Eil, David and Justin M. Rao (2011) "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 3 (2), 114–138, 10.1257/mic.3.2.114.

Eliaz, Kfir and Andrew Schotter (2010) "Paying for confidence: An experimental study of the demand for non-instrumental information," *Games and Economic Behavior*, 70 (2), 304–324, 10.1016/j.geb.2010.01.006.

Gigerenzer, Gerd, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, and Steven Woloshin (2007) "Helping Doctors and Patients Make Sense of Health Statistics," *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 8 (2), 53–96, 10.1111/j.1539-6053.2008.00033.x.

Grether, David M. (1992) "Testing bayes rule and the representativeness heuristic: Some experimental evidence," *Journal of Economic Behavior & Organization*, 17 (1), 31–57, 10.1016/0167-2681(92)90078-P.

Hammerton, M. (1973) "A case of radical probability estimation," *Journal of Experimental Psychology*, 101 (2), 252–254, 10.1037/h0035224, Place: US Publisher: American Psychological Association.

Hoffman, Mitchell (2016) "How is Information Valued? Evidence from Framed Field Experiments," *The Economic Journal*, 126 (595), 1884–1911, 10.1111/ecoj.12401.

Holt, Charles A. and Angela M. Smith (2009) "An update on Bayesian updating," *Journal of Economic Behavior & Organization*, 69 (2), 125–134, 10.1016/j.jebo.2007.08.013.

Howard, Kirsten and Glenn Salkeld (2009) "Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer," *Value in Health*, 12 (2), 354–363, 10.1111/j.1524-4733.2008.00417.x.

Kahneman, Daniel and Amos Tversky (1973) "On the psychology of prediction," *Psychological Review*, 80 (4), 237–251, 10.1037/h0034747, Place: US Publisher: American Psychological Association.

Karni, Edi (2009) "A Mechanism for Eliciting Probabilities," *Econometrica*, 77 (2), 603–606, `https://www.jstor.org/stable/40263877`, Publisher: [Wiley, The Econometric Society].

Laury, Susan K., Melayne Morgan McInnes, and J. Todd Swarthout (2009) "Insurance decisions for low-probability losses," *Journal of Risk and Uncertainty*, 39 (1), 17–44, 10.1007/s11166-009-9072-2.

Liang, Wenchi, William F. Lawrence, Caroline B. Burnett, Yi-Ting Hwang, Matthew Freedman, Bruce J. Trock, Jeanne S. Mandelblatt, and Marc E. Lippman (2003) "Acceptability of diagnostic tests for breast cancer," *Breast Cancer Research and Treatment*, 79 (2), 199–206, 10.1023/a:1023914612152.

Masatlioglu, Yusufcan, A. Yesim Orhun, and Collin Raymond (2017) "Intrinsic Information Preferences and Skewness," September, 10.2139/ssrn.3232350.

Neumann, Peter J., Joshua T. Cohen, James K. Hammitt, Thomas W. Concannon, Hannah R. Auerbach, Chihui Fang, and David M. Kent (2012) "Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents," *Health Economics*, 21 (3), 238–251, 10.1002/hec.1704.

Schwartz, Lisa M., Steven Woloshin, Floyd J. Fowler, and H. Gilbert Welch (2004) "Enthusiasm for cancer screening in the United States," *JAMA*, 291 (1), 71–78, 10.1001/jama.291.1.71.

Volkman-Wise, Jacqueline (2015) "Representativeness and managing catastrophe risk," *Journal of Risk and Uncertainty*, 51 (3), 267–290, 10.1007/s11166-015-9230-7.

Xu, Yan (2020) "Revealed Preferences over Experts and Quacks and Failures of Contingent Reasoning."

A Tables

Table 15: Demographic Characteristics of Subjects

	All		$p \in \{0.1, 0.3\}$		$p \in \{0.2, 0.5\}$	
	N	%	N	%	N	%
Male	43	41	22	41	21	41
Age>23yrs old	14	13	6	11	8	16
Students	88	84	46	85	42	82
Had statistics classes	63	60	37	69	26	51
Total	105	100	54	100	51	100

Table 16: Risk Aversion Measurement

Switching Probability (π^*)	θ	N
Always protect	>2	1
0.1	2	10
0.15	1.216	13
0.2	0.573	29
0.25	0	16
0.3	-0.539	15
Never protect	<-0.539	14

Table 21: WTP: extra effect of prior probability

	(1)	(2)	(3)	(4)
model				
FP rate	-4.43	-5.88*	-4.76*	-6.17*
	(2.8)	(3.3)	(2.8)	(3.3)
FN rate	-2.35**	-1	-2.7*	-1.46
	(1.2)	(1.6)	(1.4)	(1.8)
Stat. class			-.441*	-.436*
			(0.2)	(0.2)
Stat. class × FP rate			.809	.762
			(1.1)	(1.1)
Stat. class × FN rate			.568	.609
			(1.1)	(1.1)
Constant	1.46***	1.25***	1.77***	1.55***
	(0.2)	(0.4)	(0.3)	(0.4)
sigma				
Constant	1.88***	1.88***	1.87***	1.87***
	(0.1)	(0.1)	(0.1)	(0.1)
With squares	No	Yes	No	Yes
Observations	630	630	630	630
Adjusted R^2				

Controlling for priors and total probabilities of false-posiive and false-negative outcomes. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22: Belief updating: evidence of signal and base rate neglect

	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Prior	.238***	0	.173***	0	.113*	0
	(5.4)	(.)	(3.1)	(.)	(1.8)	(.)
Signal	.512***	.512***	.382***	.382***	.649***	.649***
	(6.7)	(6.7)	(3.6)	(3.6)	(5.8)	(5.8)
Good quiz × Prior			.131	0		
			(1.5)	(.)		
Good quiz × Signal			.245	.245		
			(1.6)	(1.6)		
Stat. class × Prior					.192**	0
					(2.3)	(.)
Stat. class × Signal					-.208	-.208
					(-1.4)	(-1.4)
Observations	140	140	140	140	140	140
Adjusted R^2	0.36	0.39	0.37	0.41	0.38	0.40

Decomposition works only for imperfect signals
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Figures

Figure 3: Theoretical vs actual WTP

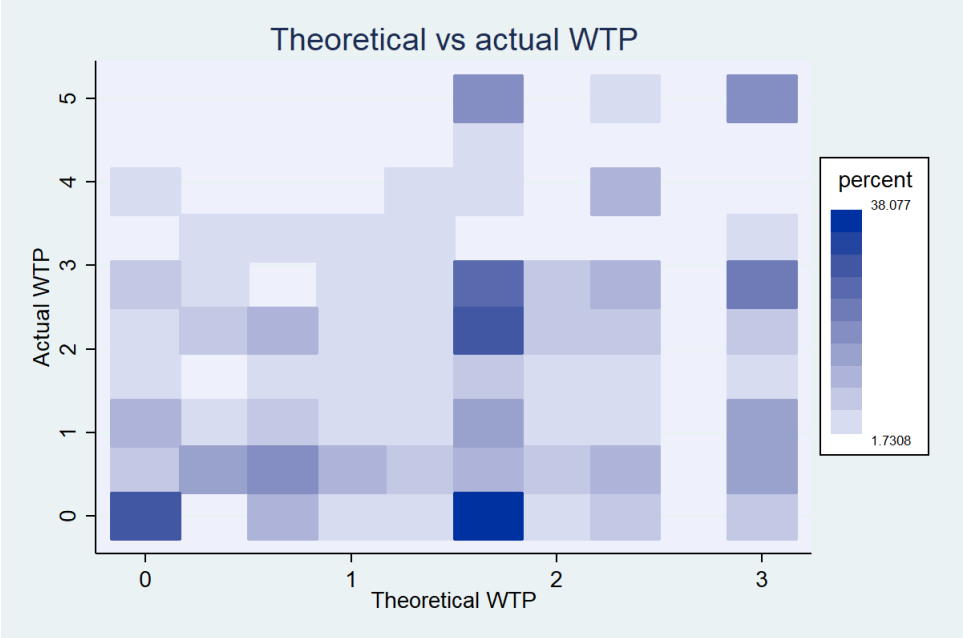
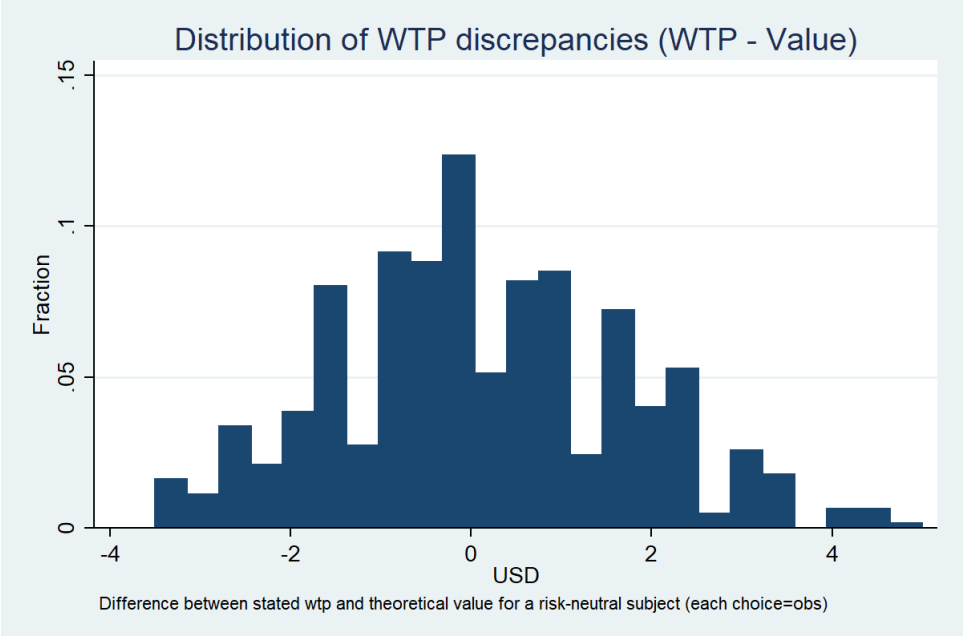


Figure 4: WTP discrepancy



C Proofs

Proof of Proposition 1:

Proof. If protection costs are low enough $c < \pi L$ than the risk-neutral decision-maker should always protect without a signal:

$$U = \max[\pi(Y - L) + (1 - \pi)Y, Y - c] = Y - c$$

It means that a strictly risk-averse decision-maker with a utility function $u()$ should also protect:

$$\pi u(Y - L) + (1 - \pi)u(Y) < u(\pi(Y - L) + (1 - \pi)Y) = u(Y - c)$$

Then denote stochastic payoff with a signal as X so that expected utility with a signal is $Eu(X-b)$ where b is the willingness-to-pay solving:

$$Eu(X - b) = u(Y - c)$$

Let b_0 be the willingness-to-pay for a risk-neutral decision-maker. By Jensen's inequality:

$$Eu(X - b_0) < u(EX - b_0) = u(Y - c) = Eu(X - b)$$

Because expected utility with a signal is a decreasing function of b_0 we obtain $b > b_0$. □

D Alternative Explanations

We show that the observed pattern of switching sensitivities to false-positive and false-negative rates with priors is not consistent with the probability-weighting and preferences for certainty.

Loss-averse decision-maker. A loss-averse decision-makers have extra-sensitivity to losses or deviations of incomes below a certain baseline. As a result, utility function becomes convex in the domain of losses. There are different functional specifications of loss aversion, but —

Probability-weighting decision-maker. A decision-maker does probability weighting if it reacts to overreacts to very low probabilities and underreacts to very high probabilities. Their behavior can be described as a standard expected utility maximization but transforming all the probabilities to number closer to the middle of the support (1/2). We can show that probability weighting implies a reverse pattern of responses to false-positive and false-negative rates with priors and hence cannot explain our observations. Willingness-to-pay solves the following equation, which is equivalent to eqation () but with probabilities x replaced by its monotonic transform $f(x)$:

$$\begin{aligned} f(P(s=1))u(Y-b-c) + f(\pi P_{01})u(Y-b-L) + f((1-\pi)P_{00})u(Y-b) = \\ = \max[u(Y-c), f(\pi)u(Y-L) + f((1-\pi))u(Y)] \end{aligned} \tag{10}$$

Taking derivatives from both sides we obtain:

$$\frac{db}{dP_{10}} = -\frac{(1-\pi)(u(Y-b) - u(Y-c-b))}{D(\pi, P_{01}, P_{10}, b)}$$

$$\frac{db}{dP_{01}} = -\frac{\pi(u(Y-c-b) - u(Y-L-b))}{D(\pi, P_{01}, P_{10}, b)}$$

Preferences for certainty. Eliaz and Schotter (2010) observe subjects paying for signals which have no potential effect on their decisions. Their theoretical explanation assumes that decision-maker's certainty in a chosen action directly affects their decision. In our setup it is equivalent to adding a strictly increasing function of the posterior belief μ to the consumption utility. Math —