Willingness-to-pay for Warnings

A. Gaduh, P. McGee and A. Ugarov

August 30, 2023

Motivation

- Many real-life situations require alarm designers to determine risk thresholds for sounding alarms: medical tests, fire alarms, extreme weather warnings, etc
- This requires balancing the costs of false-positive and false-negative events
- There is little theory and even less empirical studies on people preferences over signal characteristics

Research Questions

1. How do signal characteristics affect protective response?
2. Which signal characteristics people prefer? Two hypotheses:
   - Risk-neutral decision model provides a good description of user's preferences
   - Subjects put equal weights on costs coming from false-positive and false-negative events

Findings

- There are significant deviations from the risk-neutral model both for protective decisions and for willingness-to-pay for signals
- WTP has excess sensitivity to false-negative rates for low priors, and lower sensitivity for high priors (vice versa for false-positive)
- Subjects tend to have excessive reactions to false-positive rates with negative signals and excessive reactios to false-negative rates for positive signals (alarms)
- This effect partially comes from distorted beliefs on posterior chances of false-positive/false-negative event conditional on a signal
- This patter is not consistent with EU framework, but most consistent with decision-making heuristic in which subjects do not differentiate between false-positive and false-negative rates when choosing signals

Overview of the Experiment

Willingness to pay (WTP) for signals

- An insurance experiment:
  - Two states of the world: bad ($\omega = 1$) and good ($\omega = 0$)
  - Probability of a bad state is $P(\omega = 1) = \pi$
  - Bad state $\implies$ loss of $\$L$
  - A perfectly protective insurance can be purchased for $\$c$
- Subject can purchase a signal s before purchasing the insurance:
  - A signal is characterized by its true-positive $(P(s = 1|\omega = 1))$ and true-negative rates $(P(s = 0|\omega = 0))$

- Theoretically, what should be the WTP for a signal?
- If bad states are a priori rare ($\pi L << c$) $\implies$ never protect without a signal
- The theoretical WTP b for an expected utility maximizer given a signal s is a solution b^* to the following:

$$\begin{gathered}P(s=1)u(Y_0-b^*-c)+\pi P(s=0|\omega=1)u(Y_0-b^*-L)+\\+(1-\pi)P(s=0|\omega=0)u(Y_0-b^*)=\\=(1-\pi)u(Y_0)+\pi u(Y_0-L)\end{gathered}$$

- A risk-neutral agent would therefore pay:

$$b^*=\pi(1-P(s=0|\omega=1))L-P(s=1)c$$

Experiment Basics

1. Box with 20 white and black balls (black ball=bad state)
2. Assumptions:
   - Protection cost is \$5
   - Loss without protection is \$20
   - Cost-loss ratio is $c/L = 5/20 = 0.25$
3. Signal is an unreliable hint about the ball color
4. Vary the prior probability of bad state and the signal's information structure

Representing Signals

- A subject receives a noisy signal as a hint from one of the gremlins:

Ball/Gremlin	***Honest gremlin:*** 	***White-swamp gremlin:*** 	***Black-swamp gremlin:***
○	The Ball is white!	The Ball is white!	The Ball is black!
●	The Ball is black!	The Ball is white!	The Ball is black!

Experiment Progression

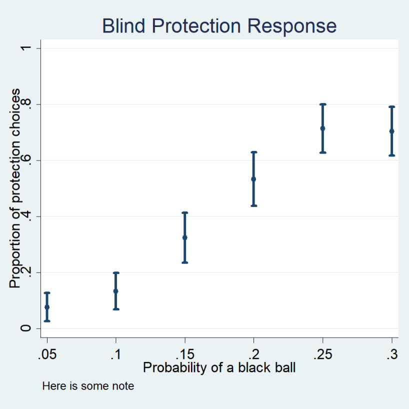
1. **Blind Protection Game:** Protection response conditional on prior probability
2. **Informed Protection Game:** Protection response conditional on prior probability and signal
3. **Belief Elicitation:** Subjects beliefs about the bad state's probability conditional on prior and signal
4. **WTP Elicitation:** Willingness-to-pay for each signal

Going over all the treatments

- Out of 4 tasks, two are novel (informed protection and WTP elicitatation for a signal)
- We report basic results to show that the subjects' behavior is largely sensible in all the tasks
- Responses demonstrate known biases in existing tasks

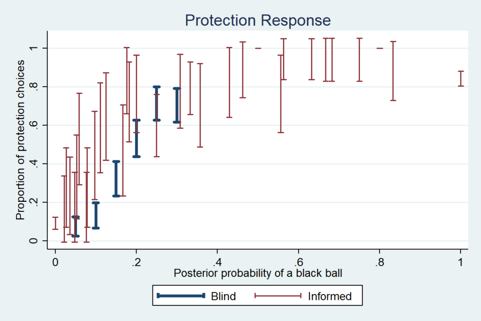
Blind Protection (No Signal)

- There is more protection when the probability of black ball is higher
- Responses are mostly monotonic in probability
- There are both risk-averse and risk-loving subjects



Informed Protection

- Protection rates increase with the posterior probability of a bad event (black ball)
- Average protection probability is no longer monotonic in posteriors due to lower N of obs per prob and due to posterior estimation mistakes



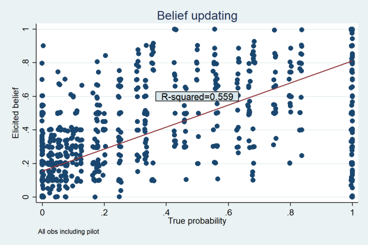
IP Big Picture

- Overprotect in response to white signals, underprotect in response to black signals without false-positive: can be explained by risk preferences
- Overprotect in response to white signals with FP (cannot be explained by risk preferences)!

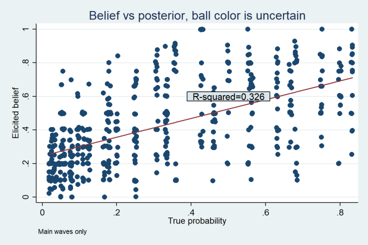
Signal	False-pos.	False-neg.	% protect	Posterior	Optimal	P(=optimal)
White	No	No	0.000	0.067	0.000	0.000
White	No	Yes	0.100	0.333	0.000	0.000
White	Yes	No	0.000	0.130	0.000	0.000
White	Yes	Yes	0.131	0.564	0.121	0.000
Black	No	No	1.000	0.846	1.000	0.000
Black	No	Yes	1.000	0.841	1.000	0.000
Black	Yes	No	0.550	0.833	0.870	0.355
Black	Yes	Yes	0.483	0.886	0.871	0.685

Belief Elicitation

- Correlation between beliefs and posteriors
- Large dispersion of errors
- Includes surprisingly large errors even for certain signals

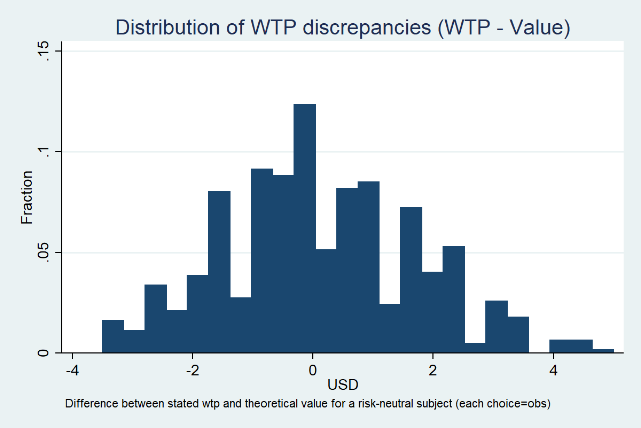

Belief updating

All obs including pilot


Belief vs posterior, ball color is uncertain

Main waves only

Distribution of WTP Discrepancies



WTP for Signals: Big Picture

- Overpaying for signals when both false-positive and false-negative events are possible
- Small and insignificant discrepancies for other signal types

False-positive	**False-negative**	**Mean WTP discrepancy**	**P$(=0)$**
No	No	-0.106	0.433
No	Yes	0.143	0.250
Yes	No	0.081	0.502
Yes	Yes	0.492	0.000

WTP for Signals

- Next, we test main hypotheses of this study
- Calculate the difference between reported WTP and theoretical WTP
- Regress the difference on FP costs ($(1-p)P(s=1|\omega=0)c$) and FN costs ($pP(s=0|\omega=1)L$)
- Coefficients should be zero if the theoretical model is correct

WTP for Signals: Determinants

	(1)	(2) FE	(3) FE	(4) FE	(5) FE
FP costs	.237*	.231*	.204	.448***	.412***
FN costs	.353***	.319***	.232	.337***	-.635***
Risk-loving × FP costs			.165		
Risk-averse × FP costs			-.0197		
No risk av. measure × FP costs			.0244		
Risk-loving × FN costs			.177		
Risk-averse × FN costs			.0608		
No risk av. measure × FN costs			.114		
Inaccurate beliefs × FP costs				-.197	
Inaccurate beliefs × FN costs				.309	
plevel=200 × FP costs					.162
plevel=300 × FP costs					-.417***
plevel=500 × FP costs					-.755*
plevel=200 × FN costs					.828***
plevel=300 × FN costs					.886***
plevel=500 × FN costs					1.02***
Constant	-.207	-.182**	-.184**	-.409***	.575***
Risk pref dummies	No	No	Yes	No	No
Prior dummies	No	No	No	No	Yes
Belief accuracy	No	No	No	Yes	No
Observations	624	624	624	624	624
Adjusted R^2	0.05	0.38	0.37	0.38	0.58

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

WTP Anomalies

- No evidence of relative underweighting of FP or FN signals on average ($\beta_{FP} = \beta_{FN}$)
- Our results exhibit the following anomaly:
  1. Excess sensitivity to false-negative costs for low priors; lack of sensitivity for high priors
  2. Underresponse to false-positive costs for low priors; excess sensitivity for high priors
- Next, we explore potential explanations for the anomaly

Anomaly: Excess Sensitivity to FP/FN Costs Varies by Priors

- Anomaly: we find relative overweighting of FN costs for low priors; overweighting of FP costs for high priors
- The pattern is monotonic by priors

Table: WTP - Value of Information, by prior

	(1) 0.1	(2) 0.2	(3) 0.3	(4) 0.5
FP costs	.437***	.576***	-.0356	-.346
	(0.1)	(0.2)	(0.2)	(0.3)
FN costs	-.645***	.196	.254***	.379***
	(0.2)	(0.1)	(0.1)	(0.1)
Constant	.467***	-.713***	-.877***	.677***
	(0.1)	(0.1)	(0.1)	(0.2)
Observations	159	153	159	153
Adjusted R^2	0.63	0.49	0.40	0.48

Standard errors in parentheses

Subject fixed effects are included.

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

Anomaly Reframed

- Note that the apparent heterogeneity comes from contrasting with the theoretical value!
- Excess sensitivity varies only because the benchmark theoretical sensitivity varies
- The sensitivity of WTP to FP/FN rates shows little variation by prior

	(1)	(2)	(3)	(4)
	0.1	0.2	0.3	0.5
FP rate	-2.91***	-2.08**	-4.46***	-3.25**
	(1.1)	(1.0)	(1.0)	(1.3)
FN rate	-2.48**	-2.73***	-3.7***	-3.65***
	(1.1)	(1.0)	(1.0)	(1.3)
Observations	159	153	159	153
Adjusted R^2				

Standard errors in parentheses
Tobit regression of WTP, constant omitted
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Potential Explanations

1. Risk preferences
2. Information preferences: paying for non-instrumental information
3. Biased beliefs:
    - Poor differentiation between false-positive and false-negative rates (all the dishonest gremlins are equally bad)
    - Neglecting to account for priors when evaluating the frequency of FP and FN events

Risk Preferences Explanation

- Can risk preferences within the EU framework lead to overweigting of FN costs and underweighting for FP costs for low priors and vice versa?
- This effect is possible in theory for FN costs if subjects are risk-averse and have positive prudency. The theory gives no answer for FP costs.
- Previous results indicate that risk aversion elicited in BP have little explanatory power for FP/FN sensitivity, but there are power issues
- We can test it more rigorously and see if controlling for risk preferences makes the prior-FP(FN) interactions small and insignificant
- We find that risk preferences do not explain the pattern for FN costs, but partially explain the variation for FP costs

Risk Preferences Testing

	(1)	(2)	(3)	(4)
			FE	FE
p>0.2	-.0953	-.875**	-.113	-.884***
FN costs	-.514	-.755	-.488	-.805*
p>0.2 × FN costs	.836**	1.24***	.826**	1.27***
Risk-loving × p>0.2 × FN costs	.245	-.733**	.164	-.633*
Risk-averse × p>0.2 × FN costs	.174	-.526	.125	-.498
No risk av. measure × p>0.2 × FN costs	.135	-.531	.158	-.655
FP costs	.506*	.334	.492*	.27
p>0.2 × FP costs	-.758*	-.189	-.734	-.191
Risk-loving × p>0.2 × FP costs	-.239	-.613	.204	-.459
Risk-averse × p>0.2 × FP costs	-.152	-.986	-.262	-.978
No risk av. measure × p>0.2 × FP costs	.158	-.92	-.453	-1.09
Risk-loving × p>0.2		.846		.837
Risk-averse × p>0.2		.982**		.967**
No risk av. measure × p>0.2		.771		.797
Risk-loving × FN costs		.807		.659
Risk-averse × FN costs		.507		.47
No risk av. measure × FN costs		.489		.681
Risk-loving × FP costs		-.0496		.314
Risk-averse × FP costs		.34		.317
No risk av. measure × FP costs		.638		.306
Risk pref dummies	No	Yes	No	Yes
Observations	624	624	624	624
Adjusted R^2	0.07	0.06	0.41	0.41

Paying for non-instrumental information

- Humans often put positive value on information not changing their decisions (citations)
- Signals with zero theoretical value can have positive elicited WTP
- We see subjects paying positive amounts for signals with zero theoretical value as well as paying for signals not affecting their decisions in IP task, but it is not clear if those are reasoning mistakes or genuine preferences
- Explaining the pattern would require non-instrumental part of value to have positive sensitivity to FP/FN rates for some priors and negative for others
- A priori there is no reason for the non-instrumental value to **decrease** with either FP or FN costs

Are Beliefs Biased?

- We are left with biased beliefs as the anomaly explanation
- Do choices in other tasks indicate biased beliefs?
- Choices in BE task show obvious biases:
    - Both the base-rate neglect and the signal underweighting
    - FN rates cause beliefs compression (lower difference in beliefs when getting black or white signals)
- In the IP task both FP and FN rates increase protection conditional on posterior when the signal is white but have an opposite effect when the signal is black $\implies$ makes signal less valuable

Base-rate Neglect Refresher

- A standard Bayesian agent does:

$$P(B|S) = \frac{P(S|B)P(B)}{P(S|W)P(W) + P(S|B)P(B)}$$

- More generally, consider an agent updating as a quasi-Bayesian:

$$\mu(B|S) = \frac{P(S|B)^{\alpha}P(B)\beta}{P(S|W)^{\alpha}P(W)^{\beta} + P(S|B)^{\alpha}P(B)^{\beta}}$$

- One can estimate it as:

$$\log\left(\frac{\mu(B|S)}{1-\mu(B|S)}\right) = \alpha \log\left(\frac{P(S|B)}{P(S|W)}\right) + \beta \log\left(\frac{P(B)}{P(W)}\right)$$

- Base-rate neglect is when $0 < \beta < 1$ and $0 < \alpha < 1$ is the signal underweighting

- We find both

Base-rate neglect and signal underweighting in BE data

	(1) OLS	(2) FE	(3) OLS	(4) FE	(5) OLS	(6) FE
Prior	.246***	.202***	.175***	.191**	.14**	.0403
	(5.5)	(4.0)	(3.1)	(2.5)	(2.3)	(0.6)
Signal	.43***	.43***	.327***	.327***	.539***	.539***
	(6.3)	(6.3)	(3.2)	(3.2)	(5.3)	(5.3)
Good quiz × Prior			.143*	.0207		
			(1.7)	(0.2)		
Good quiz × Signal			.193	.193		
			(1.4)	(1.4)		
Stat. class × Prior					.162*	.264***
					(1.9)	(2.8)
Stat. class × Signal					-.166	-.166
					(-1.2)	(-1.2)
Observations	280	280	280	280	280	280
Adjusted R^2	0.31	0.31	0.33	0.32	0.32	0.32

Decomposition works only for imperfect signals

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

Belief Biases by FP/FN rates

- FN rates induce positive belief bias for white signals and negative bias for black signals $\implies$ lowers the difference between beliefs when black/white balls
- FP rates induce positive bias both for white signals and for black signals $\implies$ the difference decreases for priors$\leq$0.2 and increases otherwise

Table: Belief Elicitation: When Mistakes Happen

	(1) All	(2) S=White	(3) S=Black
FP rate	.6***	.292***	.908***
	(0.1)	(0.1)	(0.1)
FN rate	.0108	.273***	-.251***
	(0.1)	(0.1)	(0.1)
Constant	-.0784***	.314***	-.47***
	(0.0)	(0.0)	(0.0)
Subject FE	Yes	Yes	Yes
Observations	1248	624	624
Adjusted R^2	0.15	0.41	0.52

Standard errors in parentheses
Dep. variable: reported belief - posterior probability
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Protection Biases in Informed Protection

- Estimate sensitivity to FP and FN rates by signal type and with flexible controls for posteriors
- Any excess sensitivity with flexible controls for posteriors $\Longrightarrow$ anomaly with respect to EU
- Find following anomalies:
  - FP rates increase protection when the signal is white; the same is true for FN rates but not statistically significant
  - Black signals increase protection conditional on posteriors
  - Adding flexible controls for beliefs reduces these biases

Protection Biases in Informed Protection: Results

	(1)	(2)	(3)	(4)
FP rate x (S=White)	.434***	.508***	.235*	.293**
	(3.2)	(3.4)	(1.8)	(2.0)
FN rate x (S=White)	.43*	.412	.0513	.0366
	(1.8)	(1.7)	(0.2)	(0.1)
FP rate x (S=Black)	-.128	-.0328	-.187	-.114
	(-0.3)	(-0.1)	(-0.4)	(-0.2)
FN rate x (S=Black)	.0434	-.067	.000394	-.0827
	(0.4)	(-0.4)	(0.0)	(-0.5)
S=Black	.342*	.382*	.196	.225
	(2.0)	(1.8)	(1.1)	(1.1)
p>0.2	.0504**	.0434*	.0299	.0251
	(2.5)	(1.9)	(1.5)	(1.1)
FP rate x (p>0.2)		-.131		-.101
		(-0.8)		(-0.6)
FN rate x (p>0.2)		.221		.165
		(1.3)		(1.1)
Subject FE	Yes	Yes	Yes	Yes
Posterior	Yes	Yes	Yes	Yes
Beliefs	No	No	Yes	Yes
Observations	1248	1248	1248	1248
Adjusted R^2	0.53	0.53	0.55	0.55

t statistics in parentheses

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

Do Subjects Differentiate between FP and FN rates in WTP treatment?

- The flatness of response to FP/FN rates with respect to priors can also be explained by subjects not distinguishing between FP and FN rates
- Priors have opposite effects on the frequency of FP/FN events and these effects can partially cancel each other
- Test for coefficients equality on (cost-weighted) FP/FN rates
- $\implies$ Cannot reject the hypothesis that the coefficients are equal
- Reservation: mixed evidence for Informed Protection

Differentiation of FP and FN Rates: Results

	(1)	(2)	(3)	(4)
			FE	FE
FN rate	-2.57***		-2.53***	
	(0.4)		(0.4)	
FP rate	-2.33***		-2.61***	
	(0.5)		(0.5)	
plevel=200 × FN rate		-2.23***		-2.18***
		(0.5)		(0.5)
plevel=300 × FN rate		-2.99***		-2.97***
		(0.6)		(0.6)
plevel=500 × FN rate		-2.81***		-2.77***
		(0.7)		(0.7)
plevel=200 × FP rate		-1.66**		-1.89***
		(0.7)		(0.7)
plevel=300 × FP rate		-3.31***		-3.63***
		(0.8)		(0.8)
plevel=500 × FP rate		-2.42***		-2.66***
		(0.9)		(0.9)
P(FP rate=FN rate)	.661	.862	.853	.91
Adjusted R^2	.16	.156	.543	.543
Observations	624	624	624	624

Standard errors in parentheses
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Differentiation of FP and FN rates: Informed Protection

- Evidence is mixed for the Informed Protection tasks
- We can reject the equality hypothesis at 5% significance for some specifications, but not for others
- Sensitivity to FP and FN rates also varies with priors
- It seems that subjects simplify their treatment/use heuristics when determining WTP but not when choosing protective actions as those are simpler
- Another option: heterogeneity with respect to strategies. However, while IP choices exhibit strategy heterogeneity, IP strategies do not predict WTP choices

Differentiation of FP and FN Rates: IP Evidence

Table: Informed protection response: linear probability regression

	(1) All	(2) S=White	(3) S=Black	(4) All	(5) S=White	(6) W=Blac
FP rate	.284***	.496***	.0725	.259**	.613***	-.0949
	(2.9)	(3.8)	(0.5)	(2.2)	(4.3)	(-0.5)
FN rate	.596***	1.21***	-.0213	.322***	.7***	-.0564
	(7.2)	(8.9)	(-0.2)	(3.0)	(4.0)	(-0.3)
p>0.2	.119***	.138***	.0994***	.0475**	.0438	.0512
	(6.7)	(5.6)	(3.7)	(2.0)	(1.3)	(1.4)
FP rate x (p>0.2)				.0508	-.233	.335*
				(0.4)	(-1.3)	(1.7)
FN rate x (p>0.2)				.548***	1.03***	.0703
				(4.0)	(4.4)	(0.4)
Constant	.759***	.57***	.948***	.794***	.617***	.972**
	(47.1)	(24.3)	(43.4)	(43.0)	(23.9)	(37.1)
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes
P(FP rate $\neq$ FN rate)	.00659	.000758	.631	.0025	.0000292	.495
Observations	1248	624	624	1248	624	624
Adjusted R^2	.	.	.	.	.	.

t statistics in parentheses

Errors are clustered by subject

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

Do Subjects Neglect to Account for Frequency of FP/FN events?

- The anomaly is also consistent with subject neglecting prior probabilities when evaluating the effects of FP and FN rates:
    - For the same FN rate $P(s = 0|\omega = 1)$, FN event is more likely when the priors are high as its probability is $pP(s = 0|\omega = 1)$ (the reverse is true for FP)
    - In contrast, experimental subjects reduce WTP with FP/FN rates but do it uniformly for all the priors
    - Priors still affect demand but only directly
- This is a reasonable heuristic to evaluate the value of information: account for priors, and correct according to false-positive and false-negative rates (these variables are immediately given and do not require extra computation)

Neglecting Frequency of FP/FN events

- We can test this explanation using WTP data in two ways:
  - Rewrite the interaction as the base rate plus interaction and test the joint significance of the interaction terms
  - Test for joint significance of the full set of flexible interactions betwen prior levels and FP/FN rates
- The first test cannot reject the null hypothesis, but the interaction terms are jointly significant (with or without FE) at 5%
- The change in explanatory power is very small and coefficient magnitudes are non-sensical (smaller for 0.3 rather than for 0.5)

Neglecting Frequency of FP/FN events

	(1)	(2)	(3)	(4)
			FE	FE
Prior_change×FP rate	-.392		-.457	
	(0.5)		(0.4)	
Prior_change×FN rate	-.0869		-.08	
	(0.1)		(0.1)	
plevel=200 × FP rate		.0497		.0668
		(0.2)		(0.2)
plevel=300 × FP rate		-.281***		-.281***
		(0.1)		(0.1)
plevel=500 × FP rate		-.103		-.0857
		(0.2)		(0.2)
plevel=200 × FN rate		-.000453		.000683
		(0.0)		(0.0)
plevel=300 × FN rate		-.0384*		-.0384*
		(0.0)		(0.0)
plevel=500 × FN rate		-.0295		-.0284
		(0.0)		(0.0)
P(beta_X=0)	.544	.0286	.53	.0276
Adjusted R^2	.158	.156	.543	.543
Observations	624	624	624	624

Standard errors in parentheses

Controlling for FP/FN rates, prior, constant omitted

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

Taking Stock on Explaining the Anomaly

- Neither risk preferences nor preferences for non-instrimental information seem to explain the anomaly
- Evidence of poor differentiation between FP and FN rates is obvious in WTP task and partially in IP task, but it is not clear if it is coming from beliefs or decision-making heuristics
- Mixed evidence on subjects neglecting the frequency of FP/FN events
- Tailored experimental tests are needed to distinguish between different explanations

Potential Further Tests

- Extra treatment: provide natural frequencies of false-positive/false-negative events before eliciting WTP
    - Example: *"out of 1000 experimental runs, roughly 250 involve a gremlin saying "Black" while the ball is white"*
- Elicit the probability of a black signal $P(S = B)$ within the belief elicitation task to see if beliefs on FP/FN events signals are biased
- Practice rounds so that subjects learn the frequencies of FP/FN events for each signal