Crying Wolf in the Lab

Arya Gaduh, Peter McGee, Alexander Ugarov*

November 30, 2022

Abstract

Keywords:

1 Introduction

A Tables

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	0.333	0
0.1,0.2,0.3,0.5	3	0	1	0	0.333
0.1,0.2,0.3,0.5	3	1	1 0	0.333	0.333
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

Table 2: 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 3: 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 4: Informed protection response: logistical regression

	(1) All	(2) S=White	(3) S=Black	(4) All	(5) S=White	(6) W=Black
FP rate	.248**	.557***	-.146	.198*	1.19***	-.38
	(2.2)	(4.8)	(-0.9)	(1.7)	(3.7)	(-0.8)
FN rate	.341***	.61***	-.025	.35***	1.26***	-.116
	(3.2)	(4.6)	(-0.2)	(3.2)	(12.8)	(-0.3)
S=Black	.454***			.473***		
	(89.2)			(98.4)		
plevel=200	.105***	.093*	.117**	0	0	0
	(2.8)	(1.9)	(2.1)	(.)	(.)	(.)
Subject FE	No	No	No	Yes	Yes	Yes
P(FP rate $\neq$ FN rate)	.524	.787	.621	.306	.855	.705
N	629	315	314	587	117	105
Pseudo R-squared	.333	.161	.0252	.522	.479	.0844
Log-likelihood	-291	-125	-152	-195	-41.2	-66.1

t statistics in parentheses

Errors are clustered by subject, average marginal treatment effects

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

Table 5: Informed Protection Response: logit with flexible control for posteriors

	(1)	(2)	(3)	(4)
FP rate	.54***	.66***	.526**	.496**
	(4.0)	(3.5)	(2.4)	(2.0)
FN rate	.129	.984***	.119	1.34***
	(1.0)	(2.6)	(0.7)	(3.4)
p≥0.2	.0642	.397***	.304***	.35***
	(1.6)	(9.4)	(7.3)	(7.1)
S=Black	-.0076	1.23*	-.0877	2.36***
	(-0.1)	(1.9)	(-0.7)	(3.4)
FP rate x (S=Black)		-1.87		-3.37***
		(-1.6)		(-2.9)
FN rate x (S=Black)		-.993**		-1.6***
		(-2.4)		(-4.0)
FP rate x (p $\geq$ 0.2)			.0312	.561*
			(0.1)	(1.7)
FN rate x (p $\geq$ 0.2)			-.0288	.549**
			(-0.1)	(2.3)
Observations	629	587	587	587
Adjusted R^2				

t statistics in parentheses

Reporting average marginal effects, subject FE, errors are clustered by subject.

With flexible controls of posterior probability

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

Table 6: 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
r2	2	1	-2.91958	1.881626	7.980388	-.3599557	1.725487	6.632253	.2198715
r3	2	2	-2.91958	6.699559	3.838407	.4707898	5.285504	-8.229022	.7801285

Table 7: IP response by class

	(1)	(2)	(3)
	All	Class 1	Class 2
S=Black	.628***	.357***	.772***
	(23.1)	(2.8)	(10.5)
FN rate*White hint	.691***	1.49***	.352***
	(4.3)	(3.5)	(3.3)
FP rate*White hint	.622***	.49	.47***
	(4.6)	(1.6)	(4.7)
FN rate*Black hint	-.0274	.0659	.0588
	(-0.2)	(0.2)	(0.4)
FP rate*Black hint	-.124	1.15***	-1.25***
	(-0.8)	(5.6)	(-4.8)
N	624	138	486
Pseudo R-squared	.347	.242	.543
Log-likelihood	-282	-62.3	-153

t statistics in parentheses

Errors are clustered by subject, average marginal treatment effects

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

Table 8: Correlates of Strategies Used

	(1)	(2)	(3)
Seek honest	.462***		
	(0.1)		
Other	.356***		
	(0.1)		
Female		.0782	
		(0.1)	
Age		-.00845	
		(0.0)	
Stat. classes		-.0674	
		(0.1)	
Accur. beliefs			.135*
			(0.1)
RA measure0			-.00705
			(0.0)
IP quiz			-.0635
			(0.0)
Constant	.433***	.975***	1.03***
	(0.1)	(0.1)	(0.2)
Observations	104	104	104
Adjusted R^2	0.15	0.02	0.01

Standard errors in parentheses

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

Table 9: 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 10: Latent Class Multinomial Choice Model Estimates

	lc_results Model	Class	Alt	Hint	False_prob	Posterior	Class share	BIC
r1	1	1	-2.558866	5.518452	-2.179902	-5.647592	1	599.1649
r1	2	1	-2.535444	1.90032	3.500951	1.732533	.2750615	581.0222
r1	2	2	-2.535444	.1317798	2.727107	8.918563	.7249385	581.0222
r1	3	1	-2.738694	1.552418	4.89195	1.063685	.2025011	587.5337
r1	3	2	-2.738694	3.413443	-.8342289	6.007274	.4550624	587.5337
r1	3	3	-2.738694	-3.203437	5.474852	16.56628	.3424365	587.5337

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 13: Average Protection by Signal Type

False-positive	**False-negative**	**Signal=Black**	**% protect**	**P(prot>0,<1)**	**Posterior**
No	No	No	0.038	0.022	0.000
No	No	Yes	0.838	0.000	1.000
No	Yes	No	0.186	0.000	0.045
No	Yes	Yes	0.786	0.000	1.000
Yes	No	No	0.143	0.001	0.000
Yes	No	Yes	0.739	0.000	0.395
Yes	Yes	No	0.429	0.000	0.062
Yes	Yes	Yes	0.829	0.000	0.328

Table 12: Belief Elicitation: When Mistakes Happen

	(1) All	(2) S=White	(3) S=Black	(4) All	(5) S=White	(6) S=Black
FN rate	.016	.39***	-.358***	.00219	.382***	-.378***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
FP rate	.919***	.318***	1.52***	.949***	.321***	1.58***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Constant	-.076***	.0414***	-.193***	-.248***	.139***	-.635***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Subject FE	No	No	No	Yes	Yes	Yes
Observations	630	315	315	630	315	315
Adjusted R^2	0.17	0.21	0.29	0.22	0.37	0.66

Standard errors in parentheses

Dep. variable: reported belief - posterior probability

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

Table 14: Average Belief Error by Signal Type

False-positive	**False-negative**	**Signal=Black**	**Belief error**	**P(= 0)**
No	No	No	0.039	0.001
No	No	Yes	-0.186	0.000
No	Yes	No	0.142	0.000
No	Yes	Yes	-0.337	0.000
Yes	No	No	0.118	0.000
Yes	No	Yes	0.173	0.000
Yes	Yes	No	0.245	0.000
Yes	Yes	Yes	0.192	0.000

Table 15: Average WTP discrepancy (WTP-Value) by Signal Type

False-positive	**False-negative**	**Mean WTP discrepancy**	**P(= 0)**
No	No	-0.148	0.420
No	Yes	-0.220	0.127
Yes	No	0.450	0.006
Yes	Yes	0.437	0.001

Table 16: 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.

Table 17: WTP for Information (tobit)

	(1)	(2)	(3)	(4)	(5)	(6)
	All	p=0.1	p=0.2	All	All	All
model						
FN costs	-.562**	-1.22**	-.682***	-.791***	-.68***	-.674**
	(0.2)	(0.6)	(0.3)	(0.2)	(0.3)	(0.3)
FP costs	-.631***	-.624**	-.519**	-.581***	-.485**	-.463**
	(0.2)	(0.2)	(0.3)	(0.2)	(0.2)	(0.2)
BP costs				.397***	.386***	.393***
				(0.1)	(0.1)	(0.1)
Belief change					.368	
					(0.3)	
Certainty						.799
						(0.8)
Constant	1.94***	1.72***	2.33***	.816**	.573	.0873
	(0.2)	(0.2)	(0.2)	(0.3)	(0.4)	(0.8)
sigma						
Constant	1.82***	1.86***	1.7***	1.78***	1.78***	1.78***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	315	162	153	315	315	315
Adjusted R^2						

Standard errors in parentheses

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

Table 18: WTP minus Value of Information (OLS)

	(1)	(2)	(3)	(4)	(5)
FP costs	.564***	.473***	.403	.502***	.435***
	(0.1)	(0.1)	(0.3)	(0.2)	(0.1)
FN costs	-.22*	.0351	-.495	.0816	-.62***
	(0.1)	(0.1)	(0.5)	(0.1)	(0.2)
Risk-loving			0		
			(.)		
Risk-averse			0		
			(.)		
No risk av. measure			0		
			(.)		
Risk-loving × FP costs			.12		
			(0.4)		
Risk-averse × FP costs			.104		
			(0.3)		
No risk av. measure × FP costs			-.142		
			(0.4)		
Risk-loving × FN costs			.744		
			(0.5)		
Risk-averse × FN costs			.552		
			(0.5)		
No risk av. measure × FN costs			.492		
			(0.5)		
Inaccurate beliefs				.0678	
				(0.2)	
Inaccurate beliefs × FP costs				.636	
				(0.8)	
Inaccurate beliefs × FN costs				.00218	
				(0.3)	
plevel=200					0
					(.)
plevel=200 × FP costs					.141
					(0.2)
plevel=200 × FN costs					.816***
					(0.2)
Constant	-.108	-.152*	-.149*	-.211	-.123
	(0.2)	(0.1)	(0.1)	(0.2)	(0.1)
Observations	315	315	315	315	315
Adjusted R^2	0.05	0.59	0.59	0.59	0.60

Standard errors in parentheses

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

Table 19: WTP for Information: heterogeneity by IP class

	(1)	(2)	(3)	(4)
	p<0.3	p<0.3	All	All
model				
FN costs	-.562**	-.699***	-.254***	-.386***
	(0.2)	(0.3)	(0.1)	(0.1)
FP costs	-.631***	-.73***	-1.03***	-1.15***
	(0.2)	(0.2)	(0.2)	(0.2)
Simpletons		-.804**		-.87***
		(0.4)		(0.3)
Simpletons × FN costs		.618		.63***
		(0.6)		(0.2)
Simpletons × FP costs		.393		.573
		(0.5)		(0.4)
Constant	1.94***	2.17***	2.34***	2.57***
	(0.2)	(0.2)	(0.1)	(0.1)
sigma				
Constant	1.82***	1.79***	1.97***	1.92***
	(0.1)	(0.1)	(0.1)	(0.1)
Observations	315	312	630	624
Adjusted R^2				

Standard errors in parentheses

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

Table 20: WTP minus Value of Information, connection to self-reported protection strategy

	(1)	(2)	(3)	(4)	(5)	(6)
	All	p=0.1	p=0.2	All	All	All
Seek honest	.923***	1.17***		1.18**		1.4**
	(0.3)	(0.4)		(0.5)		(0.6)
Other	.317	.395		.324		.594
	(0.2)	(0.4)		(0.5)		(0.5)
FN costs	-.236	-.0324	-1.09***	-.563	-.558***	.602
	(0.2)	(0.5)	(0.4)	(1.0)	(0.2)	(0.6)
FP costs	.551***	.667*	-.409**	.578	-.415**	.631
	(0.1)	(0.4)	(0.2)	(0.4)	(0.2)	(0.6)
Seek honest × FN costs		-.432		-.389		-.616
		(0.6)		(1.1)		(0.7)
Other × FN costs		-.0759		.216		-.355
		(0.6)		(1.1)		(0.7)
Seek honest × FP costs		-.179		-.222		-.155
		(0.4)		(0.5)		(0.7)
Other × FP costs		-.103		-.144		.0513
		(0.4)		(0.5)		(0.7)
Constant	-.587**	-.717**	1.84***	-.123	2.28***	-1.56***
	(0.2)	(0.3)	(0.2)	(0.4)	(0.2)	(0.5)
Observations	312	312	162	159	153	153
Adjusted R^2	0.09	0.09	0.08	0.08	0.07	0.08

Standard errors in parentheses

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

B Figures

Figure 1: Average Blind Protection Response

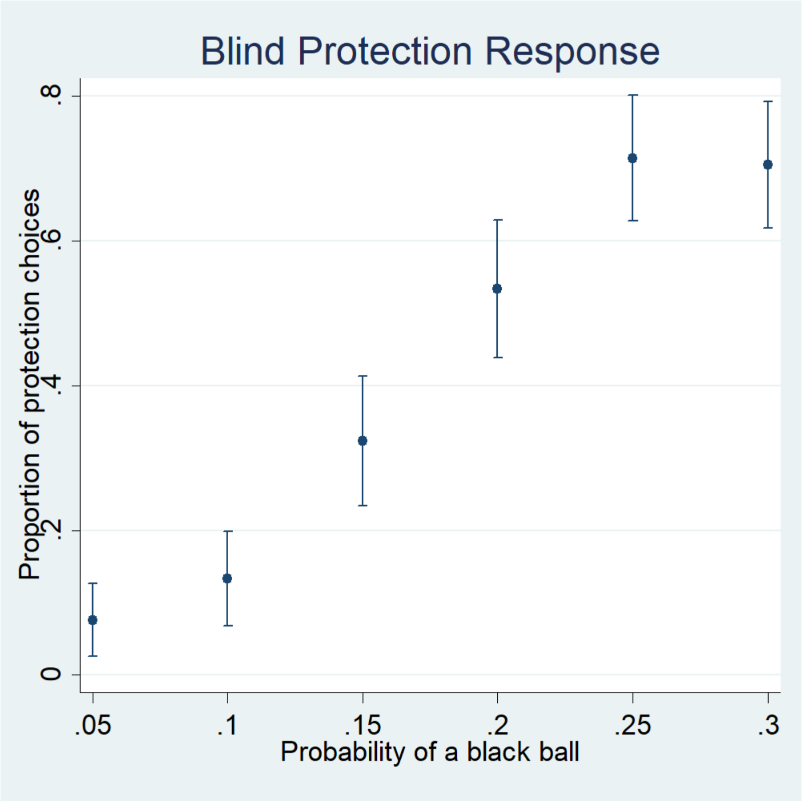


Figure 2: Average Informed Protection Response

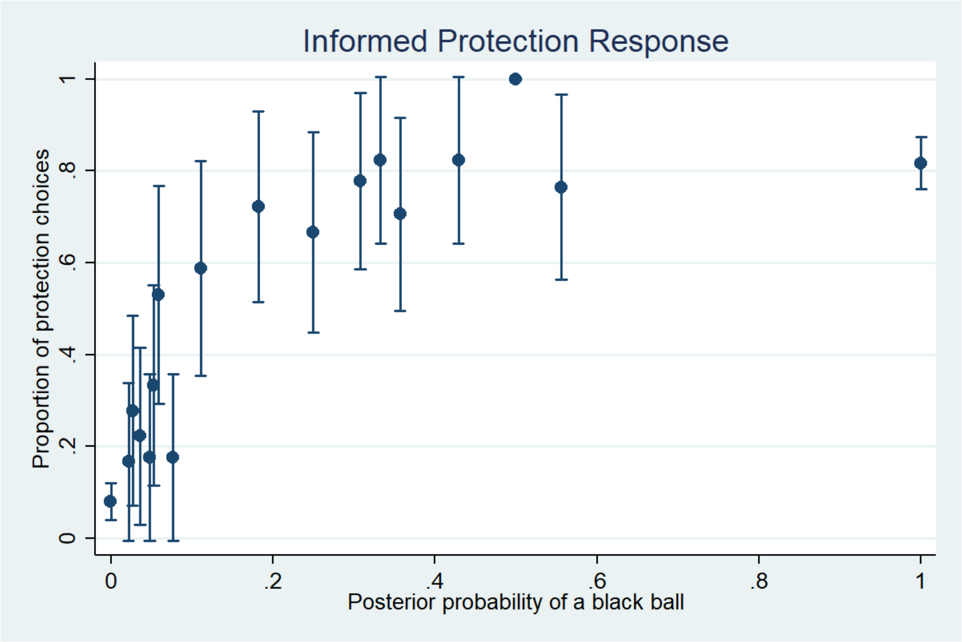


Figure 3: Average Informed Protection Response

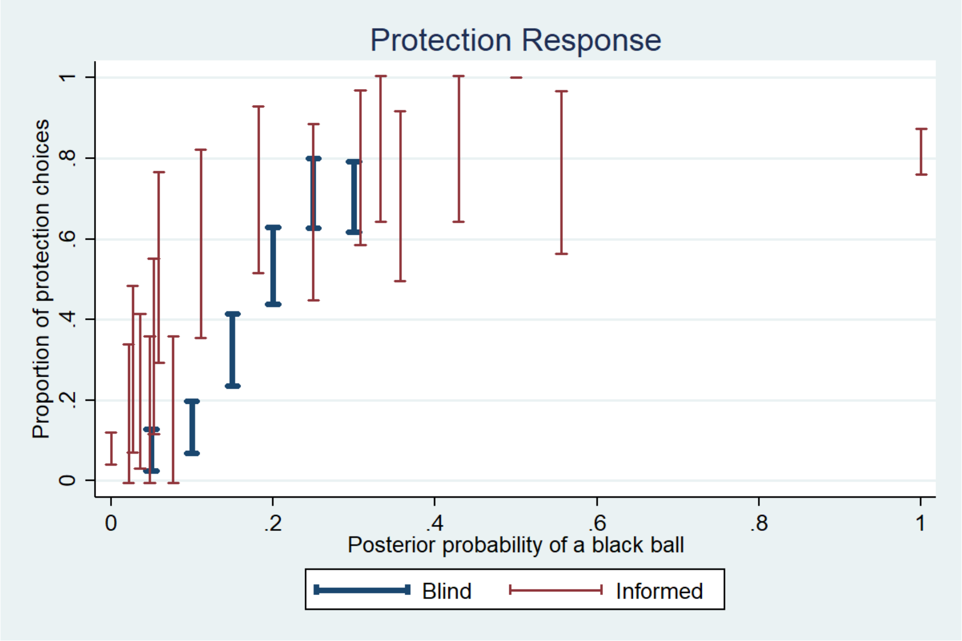


Figure 4: Average Informed Protection Response (Smoothed)

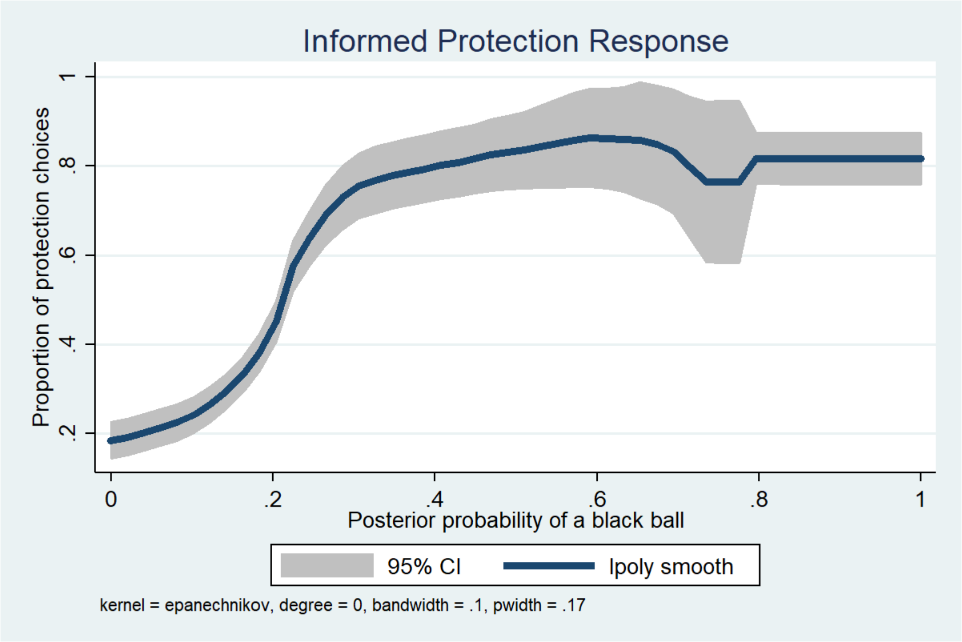


Figure 5: Belief Updating

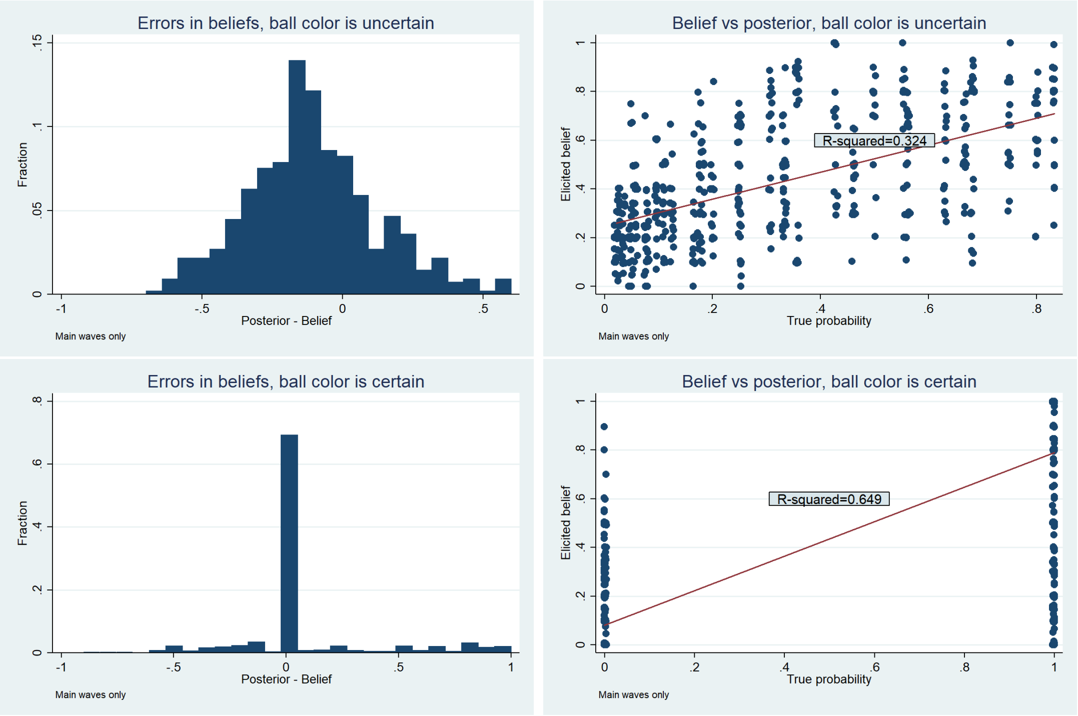


Figure 6: Theoretical vs actual WTP

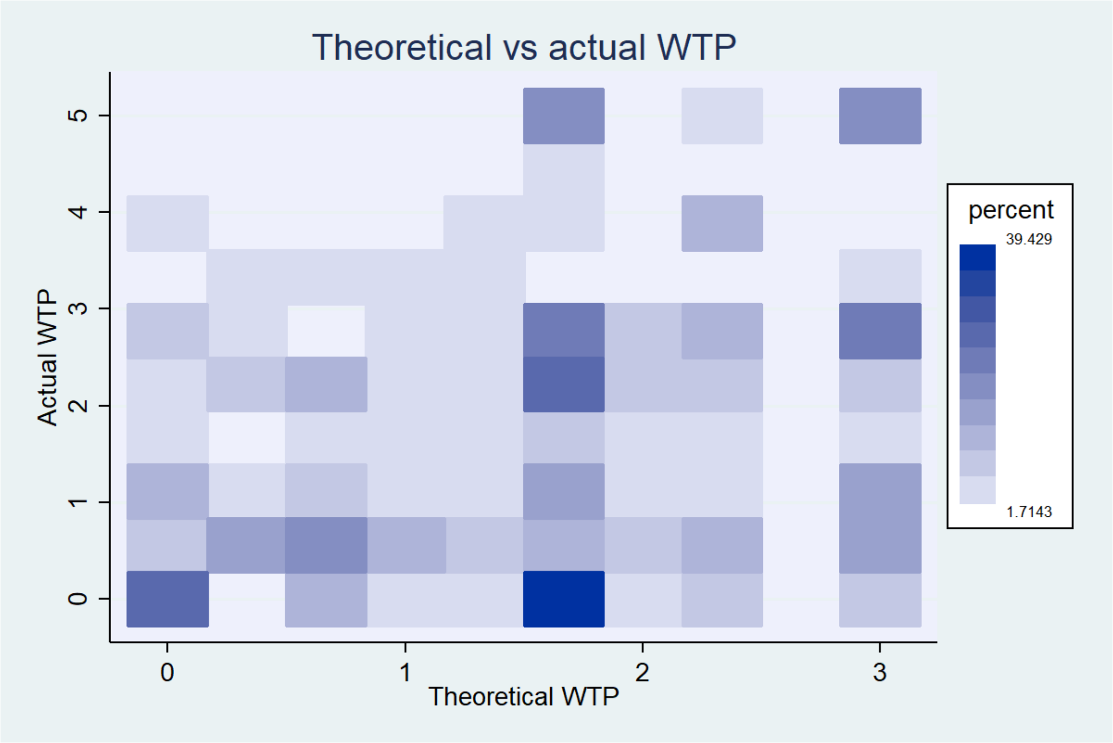


Figure 7: WTP discrepancy

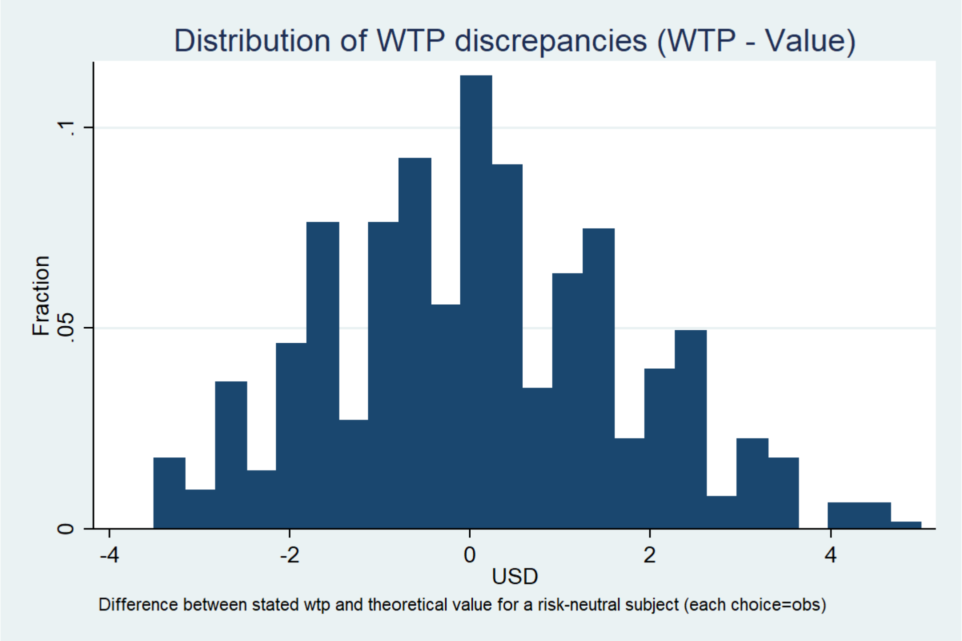
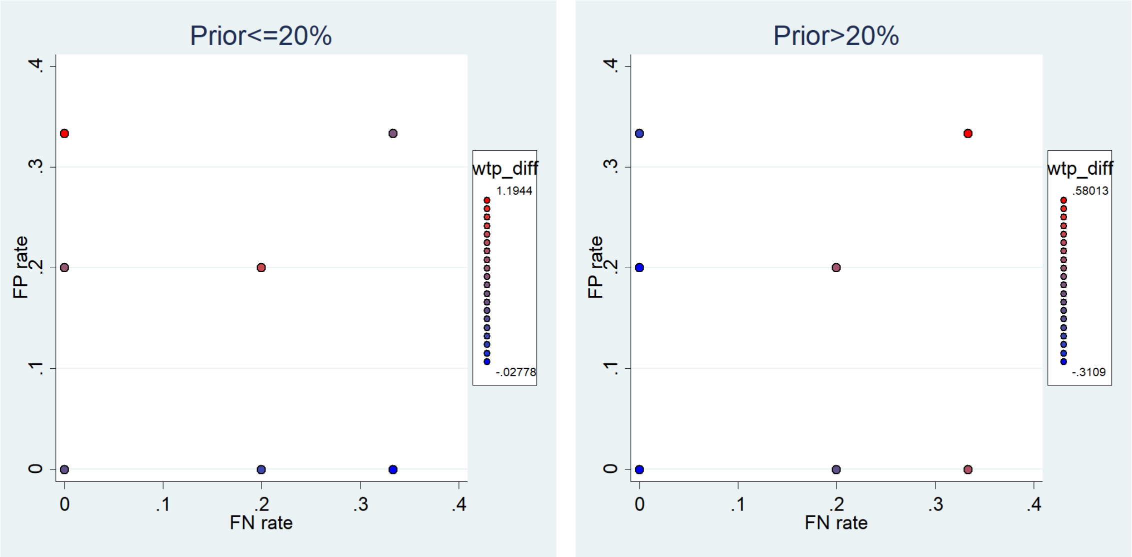


Figure 8: WTP discrepancy by prior and signal characteristics



C Appendix Tables

Table 21: Informed Protection Response: flexible control for posteriors and beliefs

	(1)	(2)	(3)	(4)	(5)	(6)
		FE			S=White	S=Black
FP rate	.329**	.292	.308	.372*	.342***	-.0805
	(2.3)	(1.5)	(1.4)	(1.9)	(2.7)	(-0.1)
FN rate	.0109	.000606	-.0916	.493	-.104	.0794
	(0.1)	(0.0)	(-0.5)	(1.3)	(-0.4)	(0.4)
p≥0.2			.279***			
			(4.7)			
FP rate x (p ≥ 0.2)			-.0175			
			(-0.1)			
FN rate x (p ≥ 0.2)			.181			
			(0.9)			
S=Black				.71		
				(1.3)		
FP rate x (S=Black)				-1.06		
				(-1.1)		
FN rate x (S=Black)				-.535		
				(-1.3)		
Observations	629	587	587	587	313	314
Adjusted R^2						

t statistics in parentheses

With flexible controls of posterior probability and beliefs

Errors are clustered by subject, average marginal treatment effects

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

Table 22: Informed protection response: semiparametric control for posteriors

	(1)	(2)	(3)	(4)
FP rate	.547***	.439**	.527***	.361*
	(3.6)	(2.2)	(3.3)	(1.8)
FN rate	-.186	-.197	-.643	.00259
	(-1.0)	(-0.9)	(-1.6)	(0.0)
p≥0.2		.0377		
		(0.8)		
FP rate x (p $\geq$ 0.2)		.225		
		(0.9)		
FN rate x (p $\geq$ 0.2)		.0451		
		(0.2)		
S=Black			-6.21	
			(-0.6)	
FP rate x (S=Black)			.00529	
			(0.0)	
FN rate x (S=Black)			.516	
			(1.3)	
Stat. class				-.0199
				(-0.4)
FP rate x Stat. class				.326
				(1.5)
FN rate x Stat. class				-.298
				(-1.4)
Observations	629	629	629	629
Adjusted R^2	0.02	0.02	0.02	0.02

t statistics in parentheses

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

Table 23: WTP - Value of Information, by prior with order effects

	(1)	(2)	(3)	(4)	(5)	(6)
	p=0.1,0.2	p=0.3,0.5	p=0.1,0.2			
FP rate	2.23***	-.249	2.12***	1.21*	-.249	-.325
	(0.5)	(0.7)	(0.7)	(0.7)	(0.7)	(0.8)
FN rate	-.254	2.64***	-1.22**	.169	2.64***	1.33***
	(0.4)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
Starts with p=0.2			-1.13***	.256		
			(0.3)	(0.3)		
Starts with p=0.2 × FP rate			.215	-.444		.157
			(1.0)	(1.0)		(0.7)
Starts with p=0.2 × FN rate			1.99***	2.11***		2.71***
			(0.7)	(0.8)		(0.6)
First prior					.0367	.0367
					(0.2)	(0.2)
First prior × FP rate					2.48***	2.48***
					(0.7)	(0.7)
First prior × FN rate					-2.9***	-2.9***
					(0.3)	(0.3)
Constant	-.135	-.172	.412*	-.278	-.172	-.172
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
Observations	315	315	315	630	630	630
Adjusted R^2	0.04	0.04	0.12	0.04	0.04	0.06

Standard errors in parentheses

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

Table 24: WTP - Value of Information, by prior

	(1)	(2)	(3)	(4)	(5)
	All	0.1	0.2	0.3	0.5
FP rate	.822*	1.96***	2.3***	-.121	-.865
	(0.5)	(0.7)	(0.7)	(0.9)	(0.9)
FN rate	1.2***	-1.24***	.783	1.57***	3.79***
	(0.4)	(0.4)	(0.5)	(0.6)	(0.7)
Constant	-.134	.435***	-.713***	-.921***	.677***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)
Observations	630	162	153	162	153
Adjusted R^2	0.36	0.64	0.49	0.42	0.48

Standard errors in parentheses

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

Table 25: Belief Elicitation: Discrepancy

	(1)	(2)	(3)	(4)	(5)	(6)
FN rate	.016	.016	-.014	-.014	-.0562	-.0554
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
FP rate	.919***	.919***	1.07***	1.07***	1.05***	1.05***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Good quiz			.0469	.0673		
			(0.0)	(0.0)		
Good quiz × FN rate			.0463	.0464		
			(0.1)	(0.1)		
Good quiz × FP rate			-.286*	-.284*		
			(0.2)	(0.2)		
Stat. class					-.00193	-.0127
					(0.0)	(0.0)
Stat. class × FN rate					.127	.126
					(0.1)	(0.1)
Stat. class × FP rate					-.229	-.226
					(0.2)	(0.2)
Constant	-.076***	-.0656***	-.101***	-.102***	-.0751***	-.0563
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Prior prob dummies	No	Yes	No	Yes	No	Yes
Observations	630	630	630	630	630	630
Adjusted R^2	0.17	0.17	0.17	0.17	0.17	0.17

Standard errors in parentheses

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

Table 26: WTP minus Value of Information: demographic determinants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
FP costs	.564***	.602***	.548***	.475**	.416**	.546***	.496***	.66***	.591***
	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)	(0.2)	(0.2)
FN costs	-.22*	-.317*	-.0684	-.242	-.0701	-.285*	-.0318	-.037	.223
	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.2)	(0.2)
Male		-.23	-.27						
		(0.4)	(0.4)						
Male × FP costs		-.126	-.131						
		(0.2)	(0.2)						
Male × FN costs		.244	.251						
		(0.3)	(0.2)						
Stat. class				-.186	-.226				
				(0.4)	(0.4)				
Stat. class × FP costs				.146	.141				
				(0.2)	(0.2)				
Stat. class × FN costs				.0344	.201				
				(0.3)	(0.2)				
>23 yrs						-.807**	-.747**		
						(0.4)	(0.3)		
>23 yrs × FP costs						.187	.148		
						(0.3)	(0.3)		
>23 yrs × FN costs						.454**	.387		
						(0.2)	(0.3)		
Good quiz								.316	.346
								(0.4)	(0.4)
Good quiz × FP costs								-.184	-.159
								(0.2)	(0.2)
Good quiz × FN costs								-.337	-.35
								(0.3)	(0.2)
Constant	-.108	-.0115	.356	.00585	.387	-.00545	.324	-.279	.0568
	(0.2)	(0.2)	(0.3)	(0.3)	(0.4)	(0.2)	(0.2)	(0.3)	(0.3)
Prior dummies	No	No	Yes	No	Yes	No	Yes	No	Yes
Observations	315	315	315	315	315	315	315	315	315
Adjusted R^2	0.05	0.04	0.11	0.04	0.12	0.06	0.12	0.04	0.11

Standard errors in parentheses

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