

Crying Wolf in the Lab

Arya Gaduh, Peter McGee, Alexander Ugarov*

December 5, 2022

Abstract

Keywords:

1 Introduction

A Tables

Table 1: List of Treatments

Prop. of black balls (p)	Gremlins composition			FP rate	FN rate
	Honest	Black-eyed	White-eyed		
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	(7) S=White	(8) W=Black
FP rate	.251** (2.2)	.556*** (4.8)	-.136 (-0.8)	.2* (1.8)	1.19*** (3.7)	-.38 (-0.8)	2.3** (2.2)	-.592 (-1.2)
FN rate	.342*** (3.2)	.615*** (4.6)	-.0304 (-0.2)	.352*** (3.1)	1.26*** (12.8)	-.116 (-0.3)	2.69*** (4.1)	-.515 (-0.9)
S=Black	.454*** (83.6)			.473*** (91.4)				
plevel=200	.106*** (2.8)	.0914* (1.9)	.12** (2.2)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
FP rate x FN rate							-6.33** (-2.4)	2.27 (1.0)
Subject FE	No	No	No	Yes	Yes	Yes	Yes	Yes
P(FP rate \neq FN rate)	.542	.766	.669	.309	.855	.705	.411	.916
N	624	312	312	582	117	105	117	105
Pseudo R-squared	.33	.159	.026	.519	.479	.0844	.56	.0925
Log-likelihood	-290	-125	-152	-194	-41.2	-66.1	-34.8	-65.5

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*** (4.0)	.658*** (3.5)	.528** (2.4)	.488** (2.0)
FN rate	.128 (1.0)	1*** (2.6)	.12 (0.7)	1.36*** (3.4)
$p \geq 0.2$.0649 (1.6)	.399*** (9.4)	.306*** (7.3)	.351*** (7.1)
S=Black	-.00664 (-0.1)	1.26* (1.9)	-.0874 (-0.7)	2.4*** (3.4)
FP rate x (S=Black)		-1.89 (-1.6)		-3.42*** (-2.9)
FN rate x (S=Black)		-1.01** (-2.4)		-1.64*** (-4.0)
FP rate x ($p \geq 0.2$)			.0292 (0.1)	.573* (1.7)
FN rate x ($p \geq 0.2$)			-.0283 (-0.1)	.556** (2.3)
Observations	624	582	582	582
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*** (23.1)	.357*** (2.8)	.772*** (10.5)
FN rate*White hint	.691*** (4.3)	1.49*** (3.5)	.352*** (3.3)
FP rate*White hint	.622*** (4.6)	.49 (1.6)	.47*** (4.7)
FN rate*Black hint	-.0274 (-0.2)	.0659 (0.2)	.0588 (0.4)
FP rate*Black hint	-.124 (-0.8)	1.15*** (5.6)	-1.25*** (-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*** (0.1)	.975*** (0.1)	1.03*** (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)	(2)
	Honesty Seekers	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-pos.	False-neg.	Signal	% protect	P(prot>0,<1)	Posterior	Optimal	P(=optimal)
No	No	White	0.038	0.022	0.000	0.000	0.045
No	No	Black	0.837	0.000	1.000	1.000	0.000
No	Yes	White	0.188	0.000	0.045	0.000	0.000
No	Yes	Black	0.783	0.000	1.000	1.000	0.000
Yes	No	White	0.145	0.001	0.000	0.000	0.001
Yes	No	Black	0.739	0.000	0.396	0.739	1.000
Yes	Yes	White	0.429	0.000	0.062	0.000	0.000
Yes	Yes	Black	0.829	0.000	0.328	0.743	0.182

Table 12: Belief Elicitation: When Mistakes Happen

	(1)	(2)	(3)	(4)	(5)	(6)
	All	S=White	S=Black	All	S=White	S=Black
FN rate	.00702 (0.1)	.38*** (0.1)	-.366*** (0.1)	.0528 (0.1)	.409*** (0.1)	-.304** (0.1)
FP rate	.948*** (0.1)	.318*** (0.1)	1.58*** (0.1)	.888*** (0.1)	.253*** (0.1)	1.52*** (0.1)
Simpletons				.28*** (0.0)	-.105*** (0.0)	.665*** (0.0)
Simpletons \times FN rate				-.177 (0.2)	-.0993 (0.2)	-.255 (0.3)
Simpletons \times FP rate				.277 (0.2)	.316 (0.3)	.238 (0.4)
Constant	-.249*** (0.0)	.139*** (0.0)	-.636*** (0.0)	-.251*** (0.0)	.14*** (0.0)	-.641*** (0.0)
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	624	312	312	624	312	312
Adjusted R^2	0.22	0.37	0.66	0.22	0.38	0.66

Standard errors in parentheses

Dep. variable: reported belief - posterior probability

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

Table 14: Average Belief Error by Signal Type

False-pos.	False-neg.	Signal	Belief error	P(= 0)
No	No	White	0.039	0.001
No	No	Black	-0.187	0.000
No	Yes	White	0.140	0.000
No	Yes	Black	-0.332	0.000
Yes	No	White	0.116	0.000
Yes	No	Black	0.177	0.000
Yes	Yes	White	0.245	0.000
Yes	Yes	Black	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	-0.135	0.465
No	Yes	-0.209	0.152
Yes	No	0.465	0.005
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	-.577** (0.2)	-1.24** (0.5)	-.682*** (0.3)	-.791*** (0.2)	-.691*** (0.2)	-.69*** (0.3)
FP costs	-.644*** (0.2)	-.647*** (0.2)	-.519** (0.3)	-.595*** (0.2)	-.508*** (0.2)	-.494** (0.2)
BP costs				.373*** (0.1)	.363*** (0.1)	.37*** (0.1)
Belief change					.332 (0.3)	
Certainty						.688 (0.8)
Constant	1.98*** (0.2)	1.79*** (0.2)	2.33*** (0.2)	.923*** (0.3)	.701* (0.4)	.293 (0.8)
sigma						
Constant	1.8*** (0.1)	1.83*** (0.1)	1.7*** (0.1)	1.77*** (0.1)	1.76*** (0.1)	1.76*** (0.1)
Observations	312	159	153	312	312	312
Adjusted R^2						

Standard errors in parentheses

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

Table 18: WTP minus Value of Information (OLS)

	(1)	(2)	(3)	(4)	(5)
FP costs	.558*** (0.1)	.472*** (0.1)	.403 (0.3)	.506*** (0.2)	.437*** (0.1)
FN costs	-.229* (0.1)	.0337 (0.1)	-.495 (0.5)	.085 (0.1)	-.645*** (0.2)
Risk-loving			0 (.)		
Risk-averse			0 (.)		
No risk av. measure			0 (.)		
Risk-loving \times FP costs			.12 (0.4)		
Risk-averse \times FP costs			.102 (0.3)		
No risk av. measure \times FP costs			-.142 (0.4)		
Risk-loving \times FN costs			.744 (0.5)		
Risk-averse \times FN costs			.549 (0.5)		
No risk av. measure \times FN costs			.492 (0.5)		
Inaccurate beliefs				.0776 (0.2)	
Inaccurate beliefs \times FP costs				.631 (0.8)	
Inaccurate beliefs \times FN costs				-.00734 (0.3)	
plevel=200					0 (.)
plevel=200 \times FP costs					.14 (0.2)
plevel=200 \times FN costs					.84*** (0.2)
Constant	-.0921 (0.2)	-.141* (0.1)	-.137 (0.1)	-.208 (0.2)	-.111 (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

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

Table 19: WTP for Information: heterogeneity by IP class

	(1) p<0.3	(2) p<0.3	(3) All	(4) All
model				
FN costs	-.577** (0.2)	-.699*** (0.3)	-.261*** (0.1)	-.386*** (0.1)
FP costs	-.644*** (0.2)	-.73*** (0.2)	-1.04*** (0.2)	-1.15*** (0.2)
Simpletons		-.804** (0.4)		-.87*** (0.3)
Simpletons \times FN costs		.618 (0.6)		.63*** (0.2)
Simpletons \times FP costs		.393 (0.5)		.573 (0.4)
Constant	1.98*** (0.2)	2.17*** (0.2)	2.39*** (0.1)	2.57*** (0.1)
sigma				
Constant	1.8*** (0.1)	1.79*** (0.1)	1.94*** (0.1)	1.92*** (0.1)
Observations	312	312	624	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) All	(2) p=0.1	(3) p=0.2	(4) All	(5) All	(6) All
Seek honest	.923*** (0.3)	1.17*** (0.4)		1.18** (0.5)		1.4** (0.6)
Other	.317 (0.2)	.395 (0.4)		.324 (0.5)		.594 (0.5)
FN costs	-.236 (0.2)	-.0324 (0.5)	-1.11*** (0.4)	-.563 (1.0)	-.558*** (0.2)	.602 (0.6)
FP costs	.551*** (0.1)	.667* (0.4)	-.424** (0.2)	.578 (0.4)	-.415** (0.2)	.631 (0.6)
Seek honest \times FN costs		-.432 (0.6)		-.389 (1.1)		-.616 (0.7)
Other \times FN costs		-.0759 (0.6)		.216 (1.1)		-.355 (0.7)
Seek honest \times FP costs		-.179 (0.4)		-.222 (0.5)		-.155 (0.7)
Other \times FP costs		-.103 (0.4)		-.144 (0.5)		.0513 (0.7)
Constant	-.587** (0.2)	-.717** (0.3)	1.88*** (0.2)	-.123 (0.4)	2.28*** (0.2)	-1.56*** (0.5)
Observations	312	312	159	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

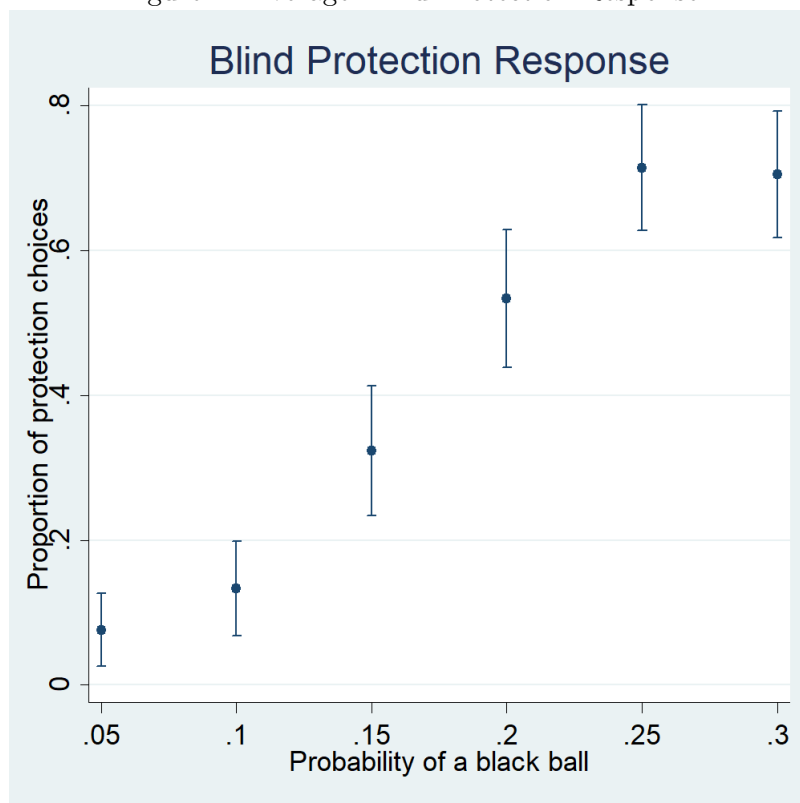


Figure 2: Average Informed Protection Response

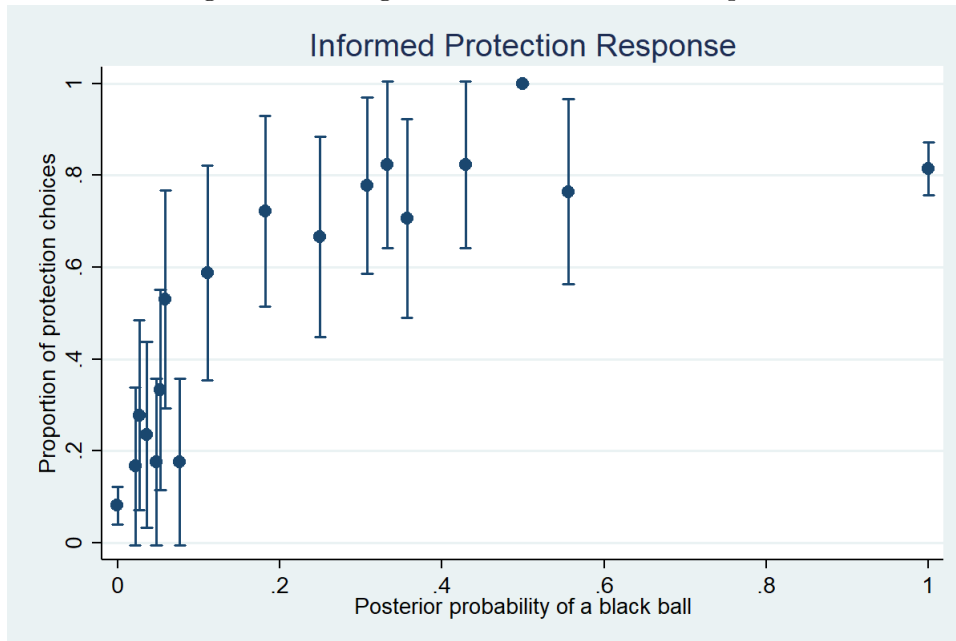


Figure 3: Average Informed Protection Response

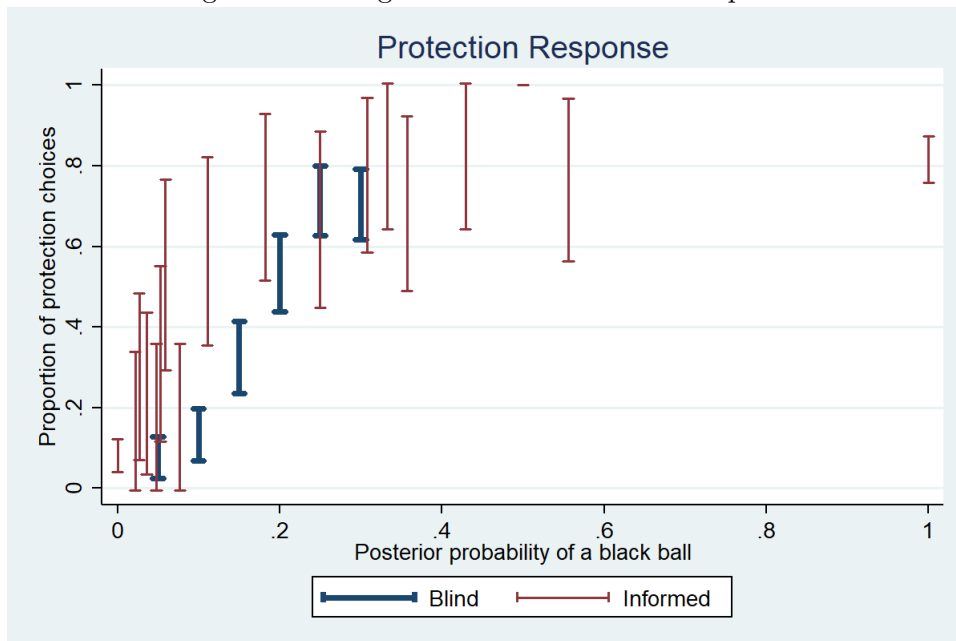


Figure 4: Average Informed Protection Response (Smoothed)

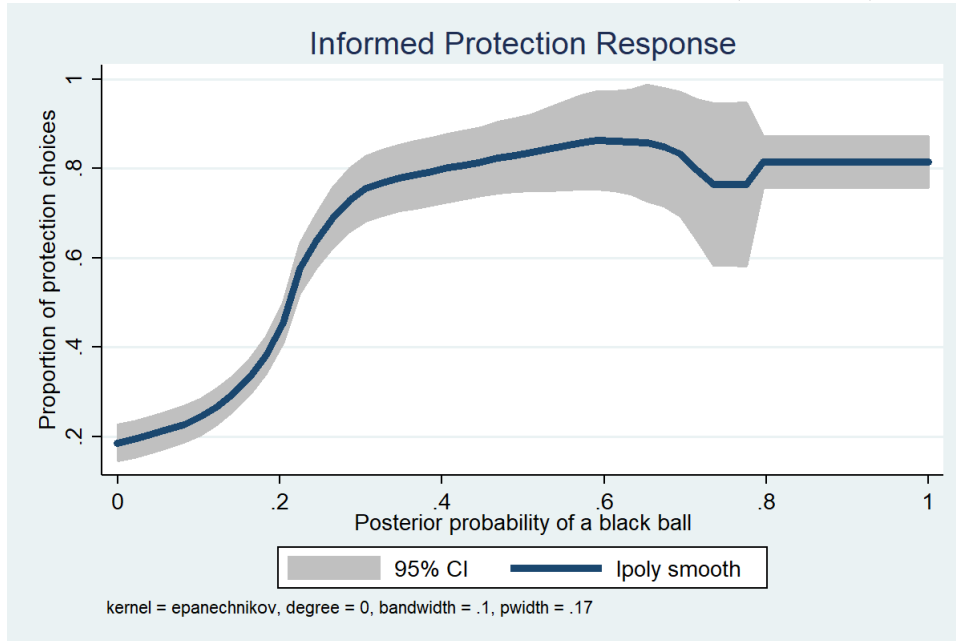


Figure 5: Belief Updating

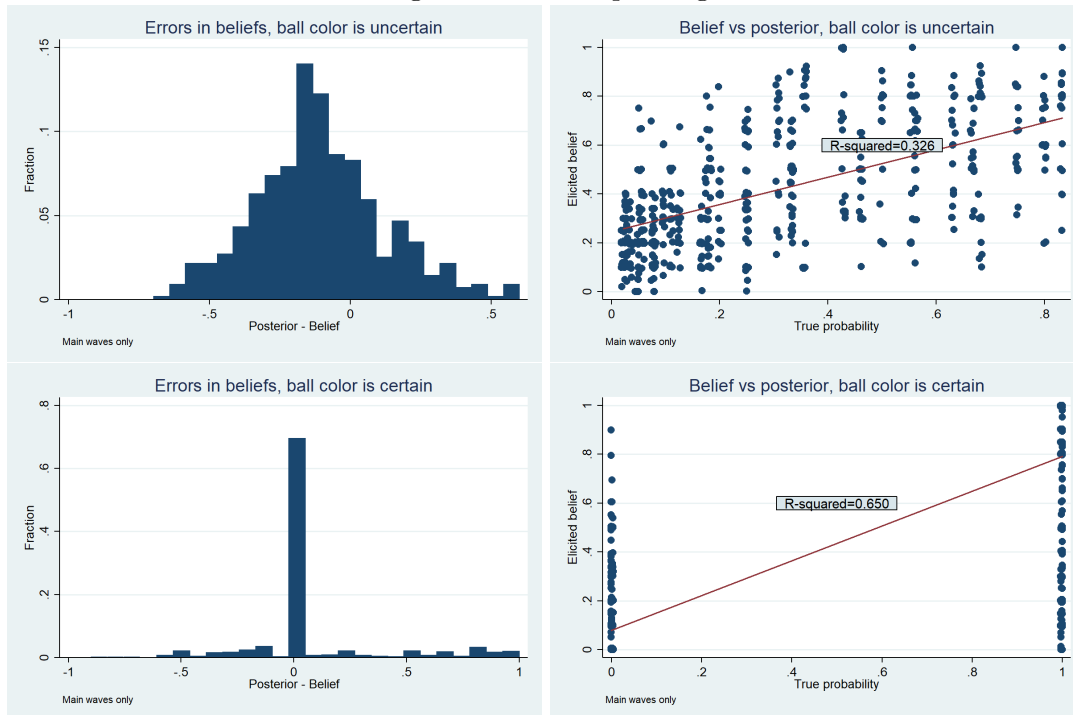


Figure 6: Theoretical vs actual WTP

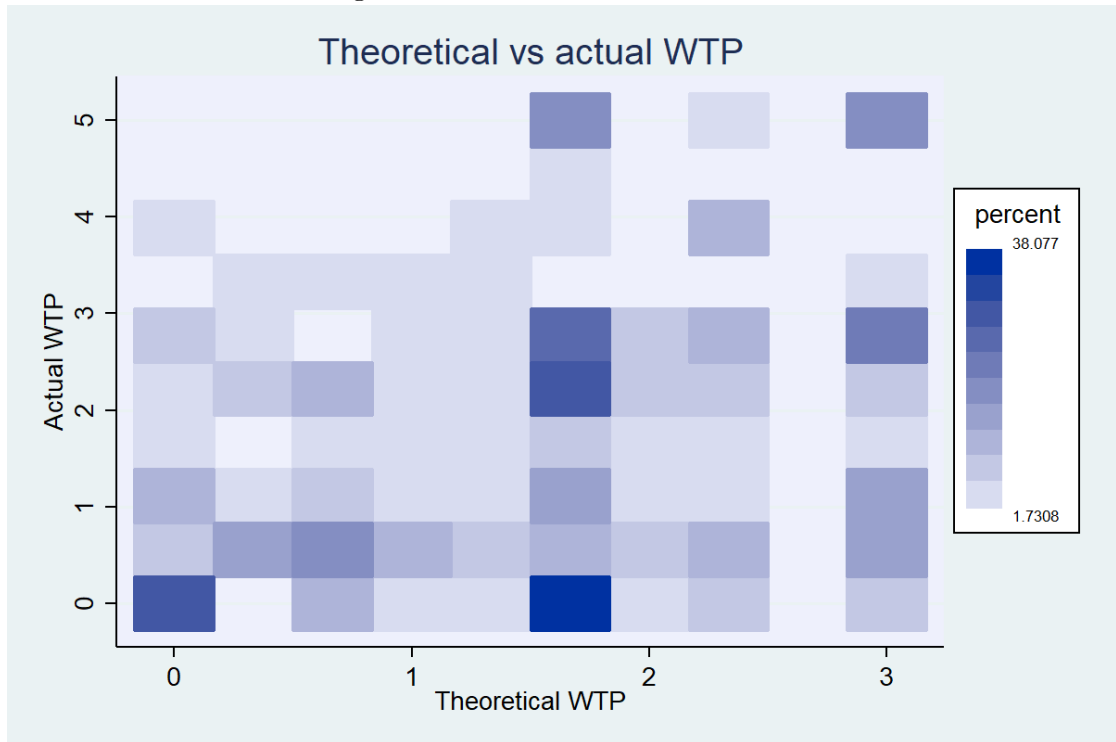


Figure 7: WTP discrepancy

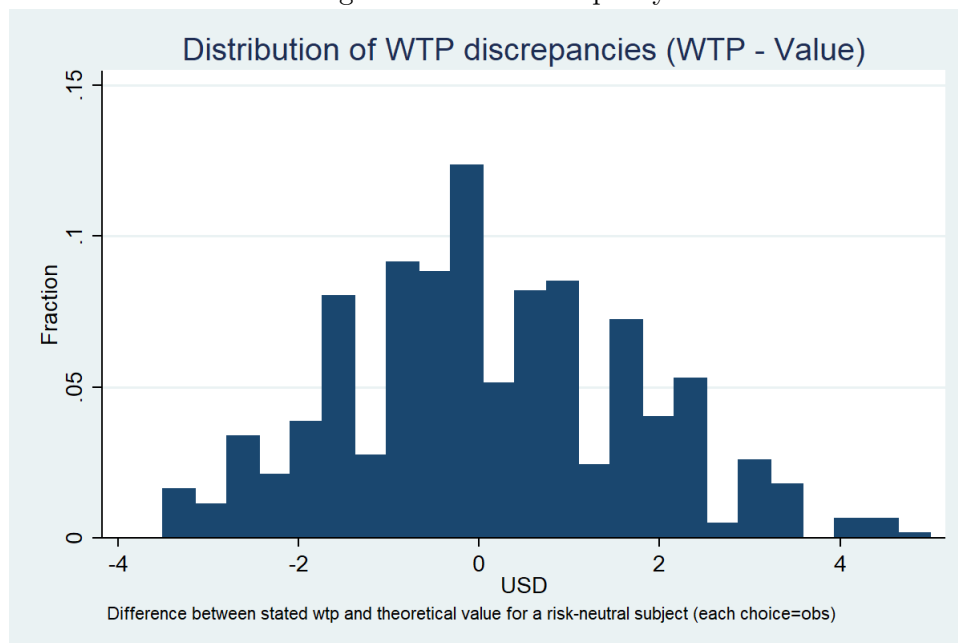
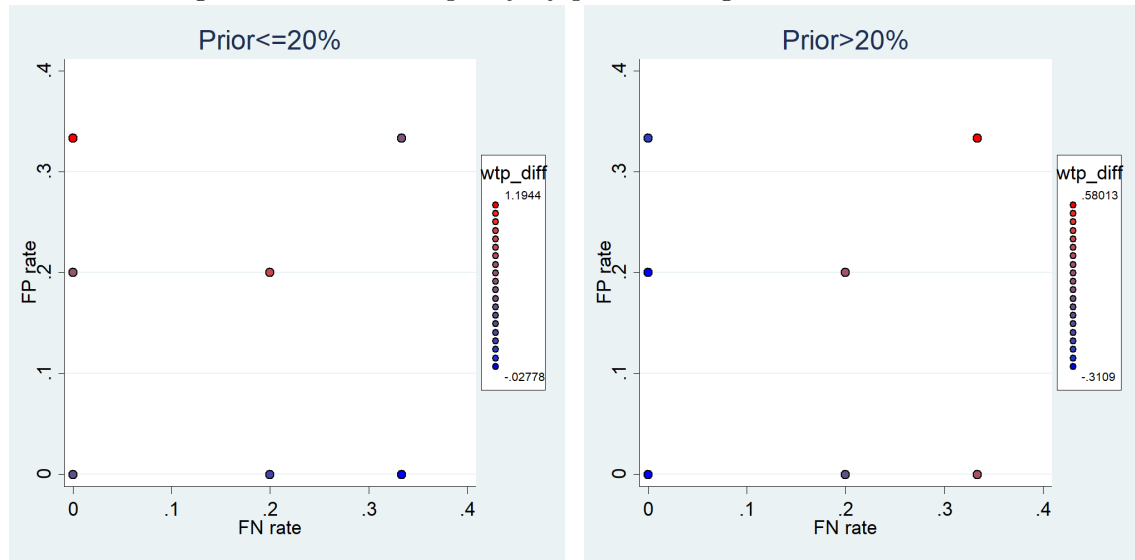


Figure 8: WTP discrepancy by prior and signal characteristics



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	.325** (2.3)	.291 (1.5)	.312 (1.4)	.369* (1.9)	.34*** (2.7)	-.0715 (-0.1)
FN rate	.00994 (0.1)	-.000178 (-0.0)	-.0956 (-0.5)	.512 (1.3)	-.0967 (-0.3)	.0767 (0.4)
$p \geq 0.2$.279*** (4.6)			
FP rate x ($p \geq 0.2$)			-.0236 (-0.1)			
FN rate x ($p \geq 0.2$)			.186 (0.9)			
S=Black				.731 (1.3)		
FP rate x (S=Black)				-1.08 (-1.1)		
FN rate x (S=Black)				-.557 (-1.4)		
Observations	624	582	582	582	310	312
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	.546*** (3.5)	.442** (2.2)	.527*** (3.3)	.357* (1.8)
FN rate	-.189 (-1.0)	-.203 (-0.9)	-.631 (-1.6)	-.000611 (-0.0)
p \geq 0.2		.0385 (0.8)		
FP rate x (p \geq 0.2)		.218 (0.9)		
FN rate x (p \geq 0.2)		.0514 (0.2)		
S=Black			-5.81 (-0.5)	
FP rate x (S=Black)			.0175 (0.0)	
FN rate x (S=Black)			.498 (1.2)	
Stat. class				-.0205 (-0.4)
FP rate x Stat. class				.333 (1.5)
FN rate x Stat. class				-.303 (-1.4)
Observations	624	624	624	624
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*** (0.5)	-.249 (0.7)	2.12*** (0.7)	1.21* (0.7)	-.249 (0.7)	-.325 (0.8)
FN rate	-.254 (0.4)	2.64*** (0.5)	-1.22** (0.5)	.169 (0.5)	2.64*** (0.5)	1.33*** (0.5)
Starts with p=0.2			-1.13*** (0.3)	.256 (0.3)		
Starts with p=0.2 \times FP rate			.215 (1.0)	-.444 (1.0)		.157 (0.7)
Starts with p=0.2 \times FN rate			1.99*** (0.7)	2.11*** (0.8)		2.71*** (0.6)
First prior					.0367 (0.2)	.0367 (0.2)
First prior \times FP rate					2.48*** (0.7)	2.48*** (0.7)
First prior \times FN rate					-2.9*** (0.3)	-2.9*** (0.3)
Constant	-.135 (0.2)	-.172 (0.2)	.412* (0.2)	-.278 (0.2)	-.172 (0.2)	-.172 (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* (0.5)	1.96*** (0.7)	2.3*** (0.7)	-.121 (0.9)	-.865 (0.9)
FN rate	1.2*** (0.4)	-1.24*** (0.4)	.783 (0.5)	1.57*** (0.6)	3.79*** (0.7)
Constant	-.134 (0.1)	.435*** (0.1)	-.713*** (0.1)	-.921*** (0.1)	.677*** (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	.021 (0.1)	.021 (0.1)	-.014 (0.1)	-.014 (0.1)	-.0562 (0.1)	-.0554 (0.1)
FP rate	.917*** (0.1)	.917*** (0.1)	1.07*** (0.1)	1.07*** (0.1)	1.05*** (0.1)	1.05*** (0.1)
Good quiz			.0467 (0.0)	.0688 (0.0)		
Good quiz \times FN rate			.0571 (0.1)	.0571 (0.1)		
Good quiz \times FP rate			-.289* (0.2)	-.288* (0.2)		
Stat. class					-.00248 (0.0)	-.0127 (0.0)
Stat. class \times FN rate					.138 (0.1)	.137 (0.1)
Stat. class \times FP rate					-.232 (0.2)	-.229 (0.2)
Constant	-.0762*** (0.0)	-.0654*** (0.0)	-.101*** (0.0)	-.102*** (0.0)	-.0751*** (0.0)	-.0563 (0.0)
Prior prob dummies	No	Yes	No	Yes	No	Yes
Observations	624	624	624	624	624	624
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	.558*** (0.1)	.602*** (0.2)	.548*** (0.2)	.475** (0.2)	.416** (0.2)	.54*** (0.1)	.485*** (0.1)	.66*** (0.2)	.591*** (0.2)
FN costs	-.229* (0.1)	-.317* (0.2)	-.0684 (0.2)	-.242 (0.2)	-.0701 (0.2)	-.295* (0.2)	-.0336 (0.1)	-.037 (0.2)	.223 (0.2)
Male		-.195 (0.4)	-.197 (0.4)						
Male \times FP costs		-.138 (0.2)	-.155 (0.2)						
Male \times FN costs		.225 (0.3)	.249 (0.2)						
Stat. class				-.161 (0.4)	-.179 (0.4)				
Stat. class \times FP costs				.138 (0.2)	.125 (0.2)				
Stat. class \times FN costs				.0192 (0.3)	.199 (0.2)				
>23 yrs						-.827** (0.4)	-.785** (0.3)		
>23 yrs \times FP costs						.193 (0.3)	.159 (0.3)		
>23 yrs \times FN costs						.465** (0.2)	.389 (0.3)		
Good quiz								.347 (0.4)	.413 (0.4)
Good quiz \times FP costs								-.194 (0.2)	-.178 (0.2)
Good quiz \times FN costs								-.355 (0.3)	-.354 (0.2)
Constant	-.0921 (0.2)	-.0115 (0.2)	.356 (0.3)	.00585 (0.3)	.387 (0.4)	.0142 (0.2)	.363 (0.2)	-.279 (0.3)	.0568 (0.3)
Prior dummies	No	No	Yes	No	Yes	No	Yes	No	Yes
Observations	312	312	312	312	312	312	312	312	312
Adjusted R^2	0.05	0.04	0.12	0.04	0.12	0.06	0.13	0.04	0.12

Standard errors in parentheses

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