Willingness-to-pay for Warnings: Pilot Results

A. Gaduh, P. McGee and A. Ugarov

April 20, 2022

Summary

- Subjects underweight both the prior probability and the signal (consistent with all the tasks using signals)
- Subjects tend to approach tasks independently
- Subjects's WTP underreact to false positive rates for low priors and overreact for high probabilities
- The opposite is true for false negative rates
- **Hypothesis: value estimation heuristics used by subjects ignores the interaction between priors and signal's characteristics**

WTP for signals: Heterogeneity with respect to priors

Table: WTP for Information (Discrepancy, by prior)

	(1)	(2)	(3)	(4)
	0.1	0.2	0.3	0.5
FP costs	.472***	.585***	.0821	-.326
	(0.2)	(0.2)	(0.2)	(0.3)
FN costs	-.609**	.192	.26**	.379***
	(0.2)	(0.1)	(0.1)	(0.1)
Constant	.412*	-.715***	-.968***	.671**
	(0.2)	(0.2)	(0.2)	(0.3)
Observations	162	153	162	153
Adjusted R^2	0.04	0.05	0.01	0.09

Standard errors in parentheses

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

WTP for signals: Heterogeneity with respect to priors

Table: WTP for Information (Discrepancy, by prior)

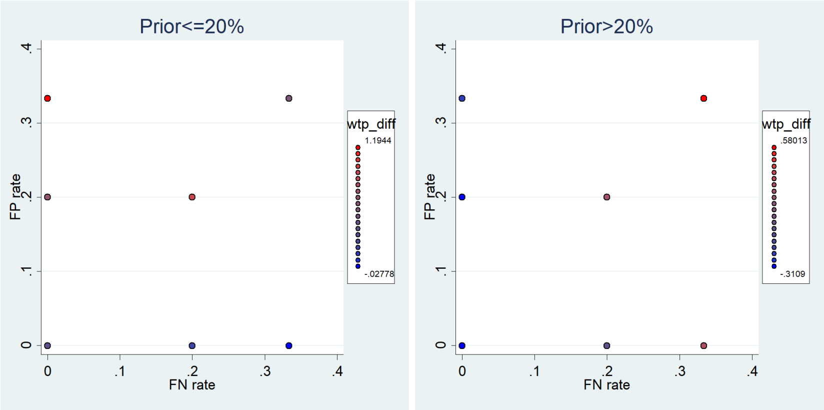
	(1)	(2)	(3)	(4)
	0.1	0.2	0.3	0.5
FP rate	2.12***	2.34***	.287	-.816
	(0.7)	(0.7)	(0.8)	(0.9)
FN rate	-1.22**	.768	1.56**	3.79***
	(0.5)	(0.5)	(0.6)	(0.7)
Constant	.412*	-.715***	-.968***	.671**
	(0.2)	(0.2)	(0.2)	(0.3)
Observations	162	153	162	153
Adjusted R^2	0.04	0.05	0.01	0.09

Standard errors in parentheses

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

WTP relative to RN agents

- Positive difference= red
- Low prior: overpaying for alerts with high FP rate
- High prior: overpaying for alerts with high FN rates



Potential Explanations for the Heterogeneity

1. Over-reacting to useless signals
2. Risk aversion with increasing third derivative (prudency)
3. Loss aversion
4. Probability weighting (part of cumulative prospect theory, rank-dependent EU, etc)
5. Probability estimation bias ("compounding neglect")

Potential Explanations: Reacting to Low-quality Signals

- The signal's value for a risk-neutral agent is bounded below at zero:

$$b^* = \max[0, \min(\pi L, c) - \pi P(s = 0|\omega = 1)L - P(s = 1)c]$$

- It is easy to prove that any signal with zero value $\pi P(s = 0|\omega = 1)L + P(s = 1)c \leq \min(\pi L, c)$ generates either too low posterior probabilities to respond to any color or too high to not respond to any color
- Potential issue: subjects are sensitive to signal's quality even for worthless signals

Potential Explanations: Reacting to Low-quality Signals

- Suppose that for given π the theoretical value of the signal is non-positive $b^* \leq 0$
- $\implies$ theoretical sensiivity to FP or FN is zero
- If WTP sensitivity is non-zero (negative) - the difference has **negative** sensitivity
- If the prior is low $(\pi L < c)$ - most bad signals have high FP, so high negative sensitivity to FP
- If the prior is high $(\pi L \geq c)$ - most bad signals come from FN, so high negative sensitivity to FN
- This prediction contradicts the observed pattern in which extra-negative sensitivity to FN emerges for low probs and vice versa
- Dropping obs with zero theoretical value or "forgetting" about bounds does not explain away the pattern

Risk-averse decision-makers

- Signal's value within EU framework is the maximum between 0 and the solution b to the following equation:

$$P(S=B)u(Y-b-c)+\pi P(S=W|\omega=B)u(Y-b-L)+$$

$$+(1-\pi)P(S=W|\omega=W)=\pi u(Y-L)+(1-\pi)u(Y)$$

  - where $P(S=W|\omega=B)$ - probability of having a negative signal (gremlin says white) conditional on the ball being black
- What can be said about WTP? Not much:
  - WTP can both increase and decrease with risk aversion even for honest treatments
  - Sensitivity to FP and FN can both increase and decrease depending on sign of $u'''()$

Risk-averse decision-makers

- Partial derivatives of WTP with respect to false-positive and false-negative rates:

$$\frac{db}{dP(B|W)} = -\frac{(1-\pi)(u(Y-b)-u(Y-c-b)}{D(\pi, P(W|B), P(B|W), b)}$$

$$\frac{db}{dP(W|B)} = -\frac{\pi(u(Y-c-b)-u(Y-L-b)}{D(\pi, P(W|B), P(B|W), b)}$$

$$D(-) \equiv P(S=H)u'(Y-c-b) + \pi P(W|B)u'(Y-L-b) +$$
$$+(1-\pi)P(W|W)u'(Y-b)$$

- Both sensitivities are negative, but we can't even say that sensitivities for str. concave $u()$ are higher or lower than risk-neutral sensitivities

- The ratio of sensitivities:

$$\frac{db/dP(B|W)}{db/dP(W|B)} = \frac{(1-\pi)}{\pi}\frac{u(Y-b)-u(Y-c-b)}{u(Y-c-b)-u(Y-L-b)} \leq \frac{(1-\pi)}{\pi}\frac{c}{L-c}$$

Risk-averse decision-makers

- For strictly concave $u()$ subjects are relatively more sensitive to false-negative rates: $|db/dP(W|B)| > |db/dP(B|W)|$
- What happens if we increase π? The first component is responsible for increasing sensitivity to false-negative rates, but there is also an effect of changing b...
- In most (interesting) cases, WTP b increases with priors π shifting arguments of utility differences
- Concavity $\implies$ difference in the numerator is "flatter", but concavity doesn't tell if the difference in flatness btw nominator and denominator goes up or down with b
- When $u'''() > 0$ (prudent) the ratio of utility differences goes up with b and hence with π $\implies$ underreacting to FN with growing π as we observe

Potential Explanations: Loss Aversion

- Many potential framework for the loss aversion, but all assume that decision-makers have convex preferences in the loss domain
- Empirically, loss aversion means overweighting small losses and underweighting large losses relative to small losses
- It requires defining the loss domain and utility functions both for gains an the losses
- Additionally, some loss aversion frameworks, including the common prospect theory also incorporate probability weighting
- Consider Gul (1991) loss aversion framework:

$$V(X) = \sum p_k u(x_k) - \beta \sum p_k I(u(x_k) < \bar{U})(\bar{U} - u(x_k))$$

- Where $\bar{U}$ is the expected utility: $\bar{U} = \sum p_k u(x_k)$
- Concave $u() \implies V()$ is convex in the loss domain

Loss Aversion

- Expected utility depends on priors, signal characteristics and utility function
- $\implies$ Losses should include false negative outcomes (worst payoff), but also can include paying for protection
- Theory has no say on whether the baseline should depend on no-protection option
- For a given threshold it is indistinguishable from the EU framework (most likely, consistent with risk aversion)
- Hence identification needs to rely on variation in the baseline outcome
- If losses=false negative, then the ratio of sensitivities

Loss Aversion

- If losses=false negative, then the ratio of sensitivities:

$$\frac{db/dP(B|W)}{db/dP(W|B)} = \frac{(1-\pi)}{\pi}\frac{(1-\beta\pi P(W|B))(u(Y-b)-u(Y-c-b))}{D}$$

$$D = [\beta P(S=B)+(1-\beta\pi P(W|B)]u(Y-c-b)+(1-\pi)\beta(1-P(B|W))u(Y-b)+$$
$$+(2\beta\pi-1-\beta)u(Y-L-b)$$

- Identical to the EU case with util. function $u()$ if $\beta = 0$
- Nominator is decreasing faster with π compared to RN case
- Denominator D can both decrease and increase depending on $u()$ curvature and parameters chosen (likely to increase for our case)
- Hence it is consistent with observed heterogen. response

Loss Aversion: no Curvature case

- If $u(x) = x$ and the loss domain includes the false-negative outcome only:

$$b = BP_c - P(S=W)(1-\beta\pi P(W|B))c - \pi P(W|B)(1+\beta-\beta\pi P(W|B))L$$

$$BP_c \equiv \min[(1+\beta(1-\pi)\pi)L, c]$$

- Sensitivity to false-positive rate decreases with π relative to risk-neutral case:

$$\frac{db}{dP(B|W)} = -(1-\pi)(1-\beta\pi P(W|B))c$$

- The effect on false-negative rate is indeterminate (depends on signal quality, cost-loss ratio):

$$\frac{db}{dP(W|B)} = -\pi((1+\beta-2\beta\pi P(L|H))L-(1-\beta(P(S=B)-\pi P(W|B)))c)$$

Alternative: Probability weighting

- In EU framework subject weight outcome by their probabilities (or their beliefs)
- In the prob. weighting framework the probabilities are rescaled towards the middle:
- Tversky and Kahneman (1992) approach rescales probability μ into:
$$\phi(\mu) = \frac{\mu^\gamma}{[\mu^\gamma + (1-\mu)^\gamma]^{1/\gamma}}, 0 < \gamma \leq 1$$
- Difference: base-rate neglect affect only probabilities not given directly, probability weighting affects all the probabilities

Probability weighting, risk-neutral case

- Use Tversky and Kahneman (1992):

$$b = \min[c, \phi(\pi)L] - \phi(P(S = B))c - \phi(\pi P(W|B))L$$

- FP response:

$$db/dP(B|W) = -(1 - \pi)\phi'(P(S = B))c$$

- $\phi'(P(S = B)) > 1$ for small $P(S = B)$, but $\phi'(P(S = B)) < 1$ for larger $P(S = B)$
- As $P(S = B)$ is increasing in π, sensitivity to FP rate decreases rel. to the baseline as π grows (consistent with obs)
- FN response:

$$db/dP(W|B) = -\pi(\phi'(\pi P(W|B))L - \phi'(P(S = B))c)$$

- For $\pi P(W|B) < P(S = B)$ the sensitivity is higher than in RN case $|db/dP(W|B)| > \pi(L - c)$ but change of sensitivity with π is unclear

Probability Estimation Bias

- Remember that WTP is:

$$b = BP_c - P(S=W)(1-\beta\pi P(W|B))c - \pi P(W|B)(1+\beta-\beta\pi P(W|B))L$$

- What if subjects do not correctly estimate the frequencies of FP $\pi P(B|W)$ and FN outcomes $(1-\pi)P(W|B)$?
- Data: subjects do not increase (decrease) sensitivity to FN (FP) rates with increasing prior
- It would explain the observed pattern
- Similar but different from the base-rate neglect theory:
  - Base-rate neglect: subjects underweight the base rate when calculating posterior probabilities
  - Here: subjects underweight the base rate when pricing signals (not accounting for growing impact of FN rates with increasing base rate)

Comparative Analysis

World	Low π FP	Low π FN	High π FP	High π FN	Ratio FP/FN
Observations	$<$	$>$	$>=$	$<$	$\uparrow$
Risk-averse and prudent	$><$	$><$	$><$	$><$	$\uparrow$
Loss-averse	$<$	$>$	$><$	$><$	$><$
Probability-weighting	$>$	$><$	$<$	$><$	$><$
Probability estimation bias	$<$	$>$	$>$	$<$	$\uparrow$

- Can exclude probability-weighting explanation, but not loss aversion and risk aversion with prudence
- Risk aversion explanation though depends on the curvature of $u()$ and the change of sensitivity with π should be very small

Comparative Analysis II

Model	Prediction
Strict risk-aversion EU	Higher sensitivity to FN rates
Strict risk-aversion EU+prudence	Ratio of FN to FP sensitivities $\uparrow$ with π
Loss aversion	FP sensit. $\downarrow$ with π FP sensit. is lower than for risk-neutral (RN)
Probability weighting	FP sensit.$>$RN for low π FP sensit.$<$RN for high π FN sensit. is higher than RN for $\pi P(W\|B) < P(S = B) < 1/2$
Probability estimation bias	FP sensitivity decreases with π rel. to RN FN sensitivity increases with π rel. to RN Diff. WTP for treatments with eq. FP and FN frequencies

Policy Implications: Probability Estimation Bias

- Providing explicit probabilities of false positive and false negative outcomes should increase quality of decision-making (expected costs)
- Providing explicit probabilities should increase WTP for high-quality alerts
- Testable predictions:
  - Signals with identical FP and FN frequencies but different priors and signal quality have different WTP (testing it in next slides)
  - Providing explicit probabilities should align WTP with the signal's value
  - With explicit probabilities subjects reduce their expected costs in the informed protection task

Accounting for Heterogeneity: Probability estimation bias

- Next, test for the probability estimation bias
- Does sensitivity to FP and FN varies with the base rate? No
- Do priors, FP and FN rates affect WTP conditional on FP and FN frequencies (interactions)? Next table - yes.

	(1) WTP	(2) WTP(good quiz)	(3) WTP(stateduc)	(4) Value(RN)
model				
p>0.2	1.02***	.918***	.999***	1.45***
	(0.3)	(0.3)	(0.3)	(0.1)
FP rate	-2.83**	-3.23**	-2.07	-4.74***
	(1.1)	(1.4)	(1.3)	(0.3)
p>0.2 × FP rate	-.374	-.982	-.962	1.64***
	(1.3)	(1.6)	(1.6)	(0.4)
FN rate	-2.45**	-3.8***	-2.24*	-1.74***
	(1.1)	(1.3)	(1.3)	(0.3)
p>0.2 × FN rate	-.874	-.373	-.797	-3.12***
	(1.3)	(1.5)	(1.6)	(0.4)
Constant	1.72***	2.11***	1.56***	1.53***

Accounting for Heterogeneity II: Probability estimation bias

	(1) WTP	(2) WTP	(3) WTP(stat)	(4) WTP(stat)	(5) Value(RN)
model					
Prior prob.	3.73***	5.31*	3.87***	5.28	1.85***
	(0.8)	(3.0)	(1.0)	(3.6)	(0.3)
FP total prob.	1.81	5.68	1.86	2.3	-6.33***
	(3.8)	(6.6)	(4.9)	(8.1)	(1.5)
FN total prob.	-2.79	-13.8	-4.8	-8.62	-16.1***
	(3.8)	(10.1)	(4.9)	(12.4)	(1.5)
FP rate	-4.43	-5.88*	-4.14	-4.34	1.1
	(2.8)	(3.3)	(3.7)	(4.3)	(1.1)
FN rate	-2.35**	-1	-1.55	-1.11	.216
	(1.2)	(1.6)	(1.5)	(2.0)	(0.4)
p squared		-2.28		-2.26	
		(4.6)		(5.6)	
FP total prob. sq.		-8.39		-.811	
		(12.6)		(15.3)	
FN total prob. sq.		57.7		20.9	
		(49.2)		(62.0)	
Constant	1.46***	1.25***	1.25***	1.09**	2.1***
	(0.2)	(0.4)	(0.3)	(0.5)	(0.1)