Revealed preferences over experts and quacks and failures of contingent reasoning

Yan Xu

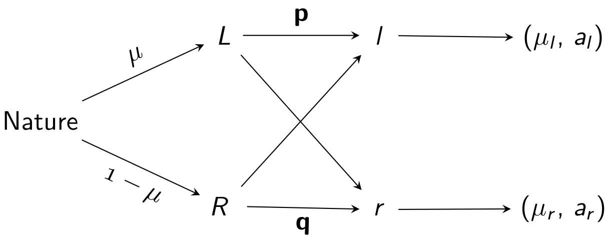
University of Vienna

July, 2022

Motivation: how do people choose and evaluate tests?

- choice set: investment advisers; doctors; medical tests ...
- decision time: before receiving a signal (advice, diagnosis)

Figure: A DM's problem of choosing a test (p, q)

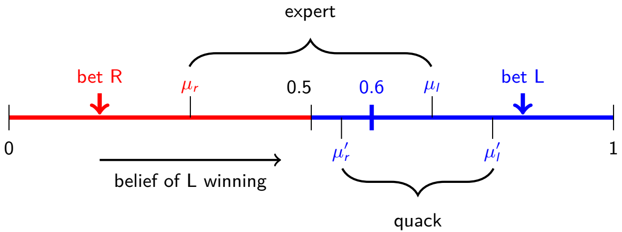


- Quacks vs. experts: useless vs. useful tests
- Can people distinguish between quacks and experts?
- What are the mechanisms of choosing quacks?

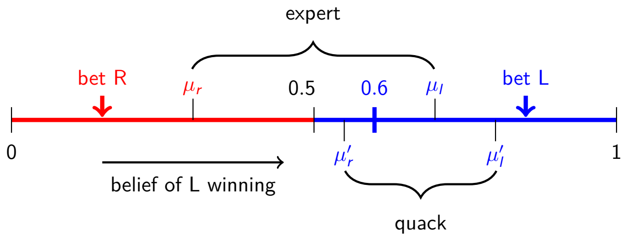
When a test (p, q) is an expert or a quack?

Task: bet state of the world (L or R) to win a prize π

- Among the previous 100 patients, L occurred 60 times and R occurred 40 times.
- Also send the patient to take a test, get a diagnosis, and then make the bet.
- A test's performance in giving correct diagnoses is:
  - Among 60 patients with tumor, it diagnosed 42 times correctly (70%)
  - Among 40 patients without tumor, it diagnosed 18 times correctly (45%)
- How much the patient should pay to get a diagnosis from this test?
  What about an alternative test whose performance is $(65\%, 55\%)$?



When a test (p, q) is a quack or an expert for a rational Bayesian DM?



Proof: Bayesian posteriors are mean preserving spreads of the prior:

$$\mu = \mathbb{E}_s \mathbb{P}(L \mid s) = \mu_l^{Bayes} s_l + \mu_r^{Bayes} s_r$$

A rational Bayesian DM's ex-ante winning probability of π is[1]:

$$v(p, q; \mu) = \begin{cases} \mu_l^{Bayes} s_l + \mu_r^{Bayes} s_r = \mu, & \text{for quacks} \\ \mu_l^{Bayes} s_l + (1 - \mu_r^{Bayes}) s_r > \mu, & \text{for experts} \end{cases}$$

[1] under structural assumptions $\mu \geq 1/2$ and $p \geq 1 - q$

Setup: states, signals, and tests

- Two states $\omega \in \{L, \ R\}$ and two signals $s \in \{l, r\}$
- The action space is binary: $u(a, \omega) = \pi \mathbb{I}_{a=\omega}$.
  — optimal action is to bet the state the DM believes $\geq 1/2$.
- The prior $\mu \equiv \mathbb{P}(\omega = L)$
- Assumption: the DM wants to maximize the chance to win the prize.
- Each test is characterized by an accuracy pair (p, q).
  — $p \equiv \mathbb{P}(s = l \mid \omega = L)$ and $q \equiv \mathbb{P}(s = r \mid \omega = R)$
- Each test induces a posterior pair (μ_r, μ_l).
  — $\mu_l(p, q; \mu) \equiv \mu(\omega = L \mid s = l)$ and $\mu_r \equiv \mu(\omega = L \mid s = r)$
- Decision scenarios: choose the most useful radiology exam, hypothesis test, statistical experiment, etc.

Mechanisms

A DM fails to distinguish between quacks and experts because he:

1. fails to **update beliefs** as a Bayesian: (μ_l, μ_r)
2. chooses **sub-optimal actions** given her beliefs: (a_l, a_r)
3. has **intrinsic preference** over certain types of tests: $skew(p, q)$
4. lacks **contingent reasoning** in the implication of a test on actions

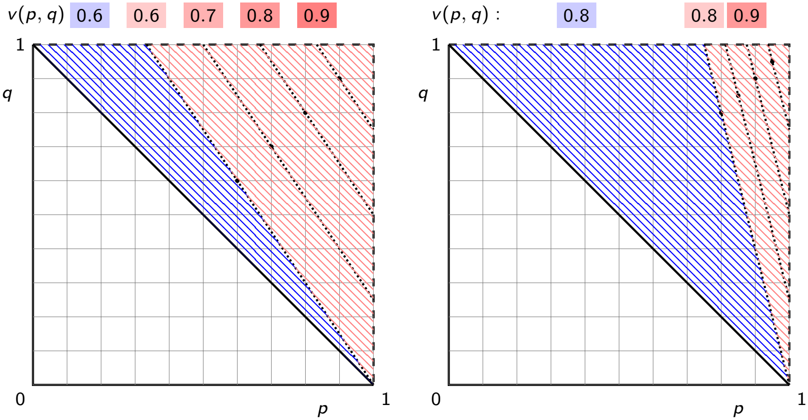
Intuition for contingent reasoning: a test is useful in providing an opportunity to contingent actions.

- quack: induced posteriors support the same optimal action (pooling): $a^*(l) = a^*(r)$
- expert: induced posteriors support different optimal actions (separating): $a^*(l) \neq a^*(r)$

This paper: elicits preferences over tests and identify different channels

Experimental design

Indifference curves of $v(p, q; \mu)$ for a rational Bayesian agent



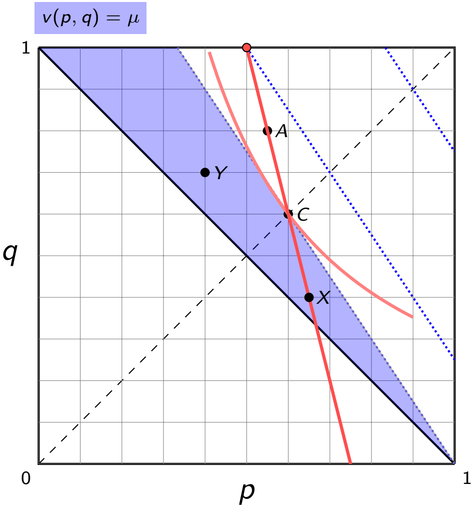
(a) Prior $\mu = 0.6$

(b) Prior $\mu = 0.8$

Figure: Value of test $v(p, q; \mu)$ for small and big priors

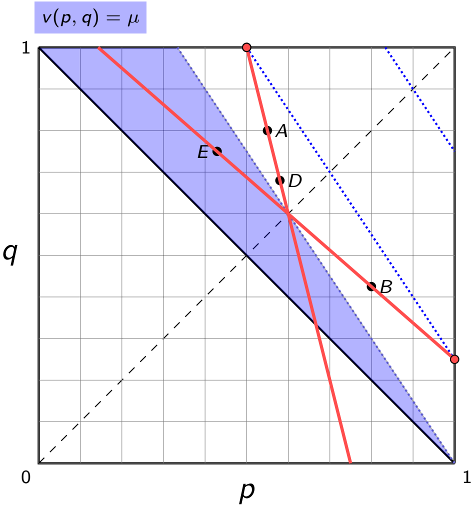
$v(p, q; \mu)$: expected winning probability of the prize for prior μ and test (p, q)

Eliciting preference over tests: trade-offs between p and q



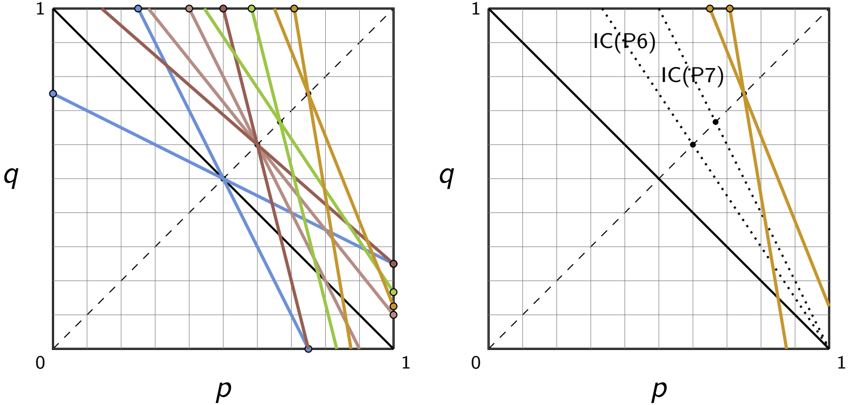
- Alternative interpretation: trade off Type I and Type II errors: $1 - p$ vs. $1 - q$
- The receiver operating characteristic (ROC) curve: p vs. $1 - q$

Eliciting preference over tests: paired linear budgets



- Budget pair: A and B are equally useful expert tests
- Identify intrinsic preference: (A, B) vs. (E, A)
- Measure the extent of intrinsic pref: (A, B) vs. $(D, B) \Rightarrow p$-skewness

Budgets for 14 rounds of tasks



(a) Budgets for P1-P5

(b) Budgets for P6-P7

Experimental task: bet state L or state R

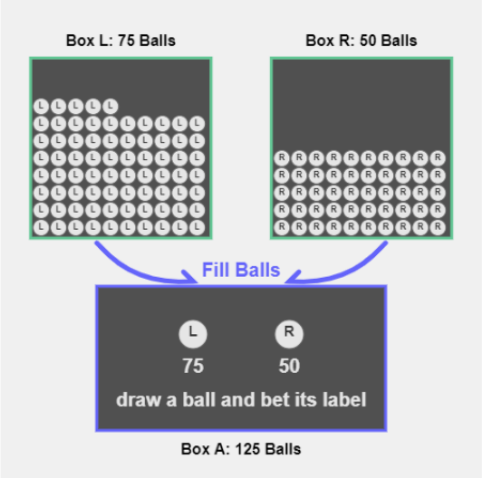
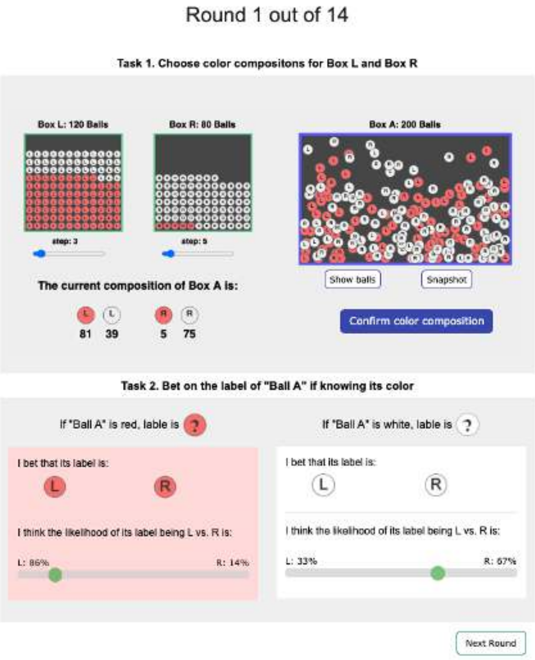
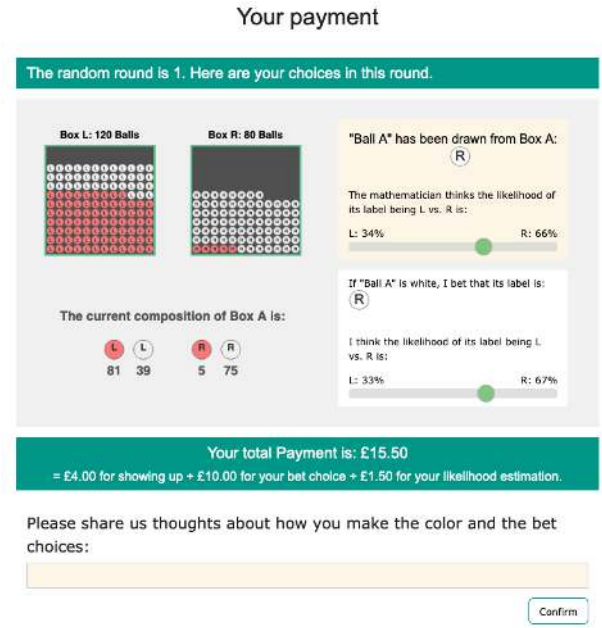


Figure: One ball (called "Ball A") will be drawn from Box A. The task is to bet its label to be either L or R. Correct bet wins a prize of £10; otherwise the payoff is 0.

Experimental task: choose a test on a budget through a coloring task



Random pay one out of fourteen rounds



Identifying different channels and experimental procedures

Identifications:

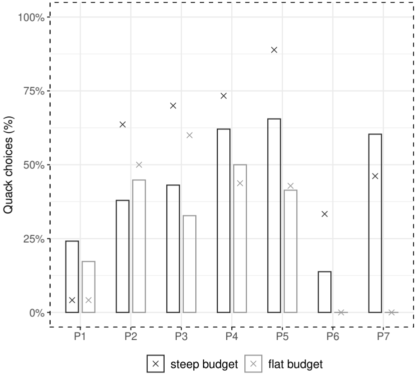
- belief-updating bias: reported posterior estimate for each signal
- best-responding bias: bet choices after each signal
- intrinsic preferences: budget pairs
- (unobservable) contingent reasoning: comments and decision rules

Procedures:

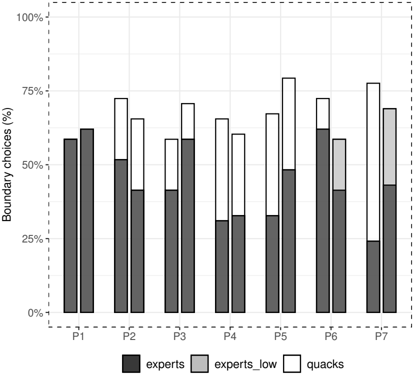
- recruit 64 (58) students on Prolific
- average payoff £11.25
- average duration 45 minutes, 18 minutes on instructions and quiz
- procedures and choices are comparable to the pilot session in the lab

Do people choose quacks?

Experimental results: failure to distinguish and evaluate quack vs. expert tests



(a) Frequency of quack choices

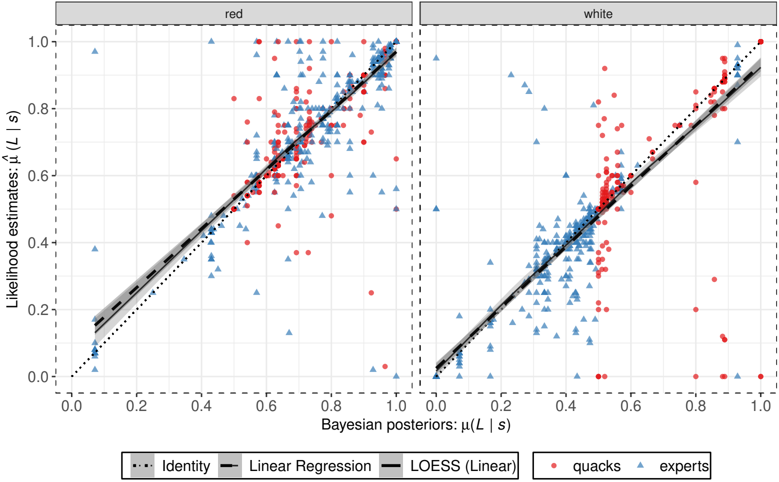


(b) Frequency of border choices

- Do people choose quacks? Yes at aggregate, round, and individual level
- What kind of tests do they choose? tests on the border
  $\Rightarrow$ the most useful experts and the most distant quacks

Channel 1: are quack choices explained by belief updating biases?

- Result 1.1: reported posteriors are close to Bayesian ones (93% bonus)



- Result 1.2: small updating biases cannot explain quack choices
- Both results are robust: OLS, IV, Grether structure regressions

Channel 2: are quack choices explained by sub-optimal actions?

Table: Number of bet choices inconsistent with the reported and Bayesian beliefs

	Under stated belief		Under Bayesian belief	
	quack	expert	quack	expert
inconsistent bets	26 1.6%	29 1.8%	35 2.2%	17 1.0%

- Result 2.1: subjects choose the optimal bets that best-respond to beliefs
- Result 2.2: small best-responding biases cannot explain quack choices

Channel 3: are quack choices explained by intrinsic preferences?

If DM cares about certain test attributes + quack tests are more likely to have the attributes $\Rightarrow$ many quacks choices

$\Rightarrow$ construct attributes measures and examine their distributions/predictability

- absolute asymmetry measures:
  - test-specific $|p-q|$, $|(p,q)-\text{pivot}|$
  - posterior-specific: $\mathbb{P}(red) = (\mu - \mu_l)/(\mu_l - \mu_r)$
- relative asymmetry measures:
  - test-specific q/p, $(q - \text{pivot})/(\text{pivot} - p)$,
  - posterior-specific: $(\mu_l - \mu)/(\mu - \mu_r)$
- All of them are similarly distributed for experts and quack tests
- None of them predicts quack choices with Probit regressions

Result 3: quacks choices cannot be justified by intrinsic preferences

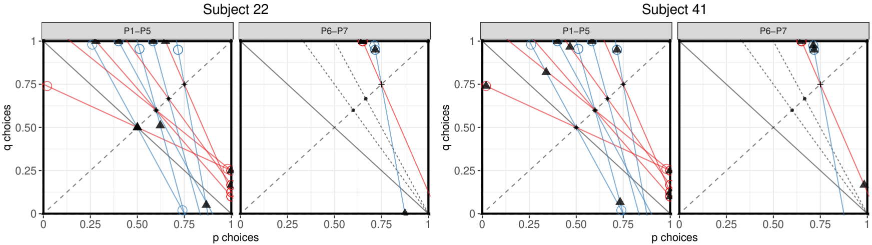
What happened?

Channel 4: are quack choices explained by the lack of contingent reasoning?

Popular decision rules describing how subjects chose coloring compositions:

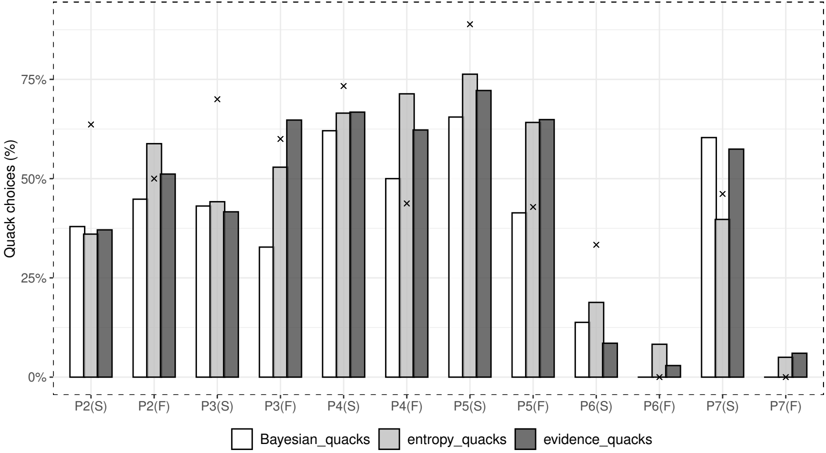
- **Entropy-reducing rule**: "I made sure that wherever I could, there was an option that red or white would 100% be label R or L"
- **Evidence-separating rule**: "The colour choices are based on the difference in red and white between L and R, you make the gap as big as possible so its easier to choose L or R from red and white."
- **Signal-separating rule**: "Try to favor one colour, increasing the chances for one colour to have a high change to belong to one of the boxes"

Figure: "I tried to somewhat increase the difference between two boxes"



Predicting the quack choice rate for each decision rule

Figure: The histogram of predicted quack choice rate for budgets in P2-P7



- two decision rules explain the choice of border tests
  – the most useful expert or most distant quack ⇒ quack choices are by-products
- simple decision rules ⇒ failure of contingent reasoning

Conclusions

- people fail to distinguish between experts and quacks
  - not because of updating bias, sub-optimal actions, or intrinsic preferences
- people over-pay for quacks but accurately pay for experts
  - because they use entropy-reducing and evidence-separating decision rules $\Rightarrow$ border tests
- people lack the contingent reasoning in choosing and evaluating tests

Contributions to the literature

- preference over information structures:
  - non-instrumental information structure:
    – timing and resolution procedure: Falk and Zimmermann (2016); Ganguly and Tasoff (2017), and Nielsen (2018)
    – skewness: Masatlioglu, Orhun, and Raymond (2017)
  - instrumentally valuable informaiton structures:
    – updating bias: Ambuehl and Li (2018)
    – prior-confirming or contradicting bias: Charness, Oprea, and Yuksel (2018); Montanari and Nunnari (2019)
  - This paper: unified framework for information structures, identifications for different channels, focus on reasoning bias

- failure of contingent reasoning:
  - violation of sure-thing principle and failure to choose dominant strategies
    – Tversky and Shafir (1992); Cason and Plott (2014); Harstad (2000); Esponda and Vespa (2014) ...
    – source of failure: not partition states (or others' action space) b/w those where DM's choice does or does not matter
  - This paper: not partition test space b/w those with which DM's optimal strategies are pooling or separating across signals.

- a tool to elicit test/source preference explicitly: rational inattention (implicitly)

Extensions and discussions

On contingent reasoning bias:

- non-binary signals and states: decision problem is not responsive
- asymmetric prize: change the threshold (elicitable)
- dynamic setting: the optimal way to acquire information at time 0?
  — e.g., lottery $(40, A_1; 15, A_2; 10, A_3)$ vs. 20, what is the optimal way to pay ϵ and ask "Is the realized state A_i or not?"
- strategic interactions: Bayesian persuasion, communication games

More questions than answers:

- How to model the failure of contingent reasoning?
- How to structure the decision rules in choosing tests?
- de-biasing: standard methods have a bound, new methods on reasoning?

Some discussions related to physician skills and credence goods

Contingent reasoning is a component in physicians' diagnostic skills:

- Variations in health expenditures are most driven by physicians' practice styles and diagnostic skills (Finkelstein et al., 2016, Cutler et al.,2019)
- Many evidence on over-testing, over-medication, and useless procedures
- Two major mistakes in diagnoses/treatments: overuse and misuse
  – C-section in childbirth: Currie et al.,2017
  – CT for pulmonary embolism(PE): Abaluck et al.,2016, Chan et al.,2021
- Empirically, does contingent reasoning bias cause overuse and misuse?
- If so, how to teach physicians to reason and make better decisions?

Credence goods: private information + mis-aligned interests

- Framework: Bayesian persuasion + private info and communication
- A doctor prefers R (regardless of states); a patient wants to match state
- The doctor commits to a test, learns the diagnosis (signal), and then communicates with the patient.
- How communication protocols affect persuasion and info transmission?

Thanks for your patience!

Comments are welcome: yan.xu@univie.ac.at [2]

[2] The working paper and slides can be found in my personal website: https://yanxu.me/

Appendix

What are the consequences of choosing quacks and non-optimal experts?

	mean	sd	pt5	pt25	pt50	pt75	pt95
Pool	5.6%	0.074	0%	0%	3.3%	8.3%	21.5%
Quack	11.6%	0.077	3.3%	6.7%	8.3%	16.7%	24.0%
Expert	2.3%	0.047	0%	0%	0%	2.5%	12.7%

Table: Relative improvements in winning probabilities if choosing optimally

Alternative definitions of expert and individual quack tests

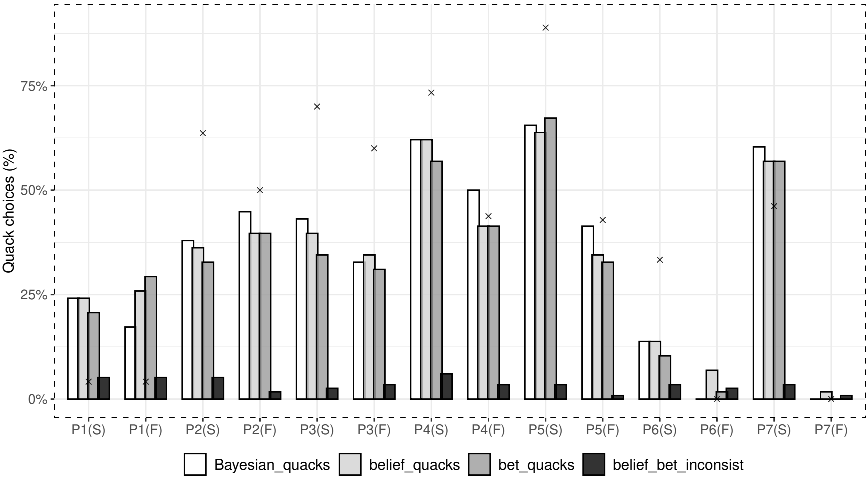


Figure: The histogram of quack choices under alternative definitions.

Predicting the quack choice rate for each decision rule

	Dependent: D(expert choice)				*Dependent:* D(top choice)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	-1.83	-45.74**	-3.46	0.13*	-5.69*	-22.66**	0.56
	(2.36)	(8.43)	(6.13)	(0.06)	(2.35)	(7.82)	(5.17)
Slope	0.84	10.60**	-1.12		1.40**	5.44**	-0.11
	(0.52)	(1.91)	(1.15)		(0.51)	(1.75)	(0.97)
Size	-1.33$\cdot$	-14.88**	0.97		-2.27**	-7.81**	-0.19
	(0.72)	(2.66)	(1.51)		(0.71)	(2.44)	(1.27)
Quack chance	-3.72**	-3.52**	-2.72**		-1.59**	-1.00**	-1.18*
	(0.52)	(0.38)	(0.69)		(0.48)	(0.30)	(0.58)
Steep	0.89*	2.48**	0.85**		1.34**	1.71**	1.01**
	(0.41)	(0.47)	(0.29)		(0.41)	(0.45)	(0.29)
Pivot point	9.32$\cdot$	104.96**	7.74		13.53**	50.36**	-0.98
	(4.99)	(18.57)	(11.79)		(4.88)	(17.05)	(9.85)
D(Top choice)				0.44**			
				(0.10)			
Top: Δ(entropy)	-5.28*				-4.50*		
	(2.22)				(2.10)		
Bottom: Δ(entropy)	-3.44				-4.38*		
	(2.19)				(2.17)		
Top: $\lvert p+q-1\rvert$		-25.89**				-11.03**	
		(4.54)				(4.10)	
Bottom: $\lvert p+q-1\rvert$		-13.58**				-7.42**	
		(2.87)				(2.60)	
Top: $\mathbb{P}$(red)			-2.30				2.03
			(3.88)				(3.20)
Bottom: $\mathbb{P}$(white)			12.55**				2.27
			(2.89)				(2.66)
Observations	696	696	696	696	696	696	696

Demographics and quack choices

	Dependent variable: individual rate of quacks					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.23	0.20	0.38***	0.14	0.03	0.24
	(0.16)	(0.19)	(0.05)	(0.14)	(0.12)	(0.18)
Age	0.004					
	(0.01)					
Female	0.04					
	(0.04)					
SAT	0.001					
	(0.02)					
STEM	0.02					
	(0.04)					
CRT score			-0.04**	-0.03*	-0.03**	-0.03**
			(0.02)	(0.02)	(0.02)	(0.02)
Wason score			-0.02			
			(0.02)			
Logic score			0.02			
			(0.02)			
Risk		0.04**		0.04***	0.04***	0.04**
		(0.02)		(0.01)	(0.01)	(0.01)
Contingent		-0.04		-0.04		-0.04
		(0.03)		(0.03)		(0.03)
Stubborn		0.02		0.03*	0.03*	0.02
		(0.02)		(0.01)	(0.01)	(0.01)
Information		0.06**		0.05**	0.04*	0.06**
		(0.03)		(0.02)	(0.02)	(0.02)
Perspective		-0.02				-0.02
		(0.03)				(0.03)
Analytical		-0.02				
		(0.02)				
Observations	58	58	58	58	58	58
Adjusted R^2	-0.06	0.15	0.04	0.21	0.19	0.21
Note:						*p<0.1; **p<0.05; ***p<0.01

Demographics and individual belief updating biases

	Dependent variable: individual coefficient $1 - \hat{\alpha_1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.03	0.11	0.20^{**}	0.12	0.22	0.07
	(0.25)	(0.38)	(0.09)	(0.20)	(0.19)	(0.21)
Age	-0.004					
	(0.01)					
Female	0.24^{***}			0.21^{***}	0.21^{***}	0.21^{***}
	(0.07)			(0.05)	(0.05)	(0.05)
SAT	0.004					
	(0.03)					
STEM	0.06					
	(0.07)					
CRT score			-0.09^{***}	-0.08^{***}	-0.07^{***}	-0.09^{***}
			(0.03)	(0.03)	(0.03)	(0.03)
Wason score			-0.04			
			(0.03)			
Logic score			0.06			0.04
			(0.04)			(0.03)
Risk aversion		0.03				
		(0.03)				
Contingent		0.01				
		(0.06)				
Stubborn		-0.02				
		(0.03)				
Information		0.03		0.06		0.05
		(0.05)		(0.04)		(0.04)
Perspective		-0.07		-0.10^{**}	-0.08^{**}	-0.09^{**}
		(0.05)		(0.04)	(0.04)	(0.04)
Analytical		0.02		0.06	0.07^{**}	0.06
		(0.05)		(0.04)	(0.04)	(0.04)
Observations	58	58	58	58	58	58
Adjusted R^2	0.17	-0.06	0.13	0.33	0.31	0.33
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$