Revealed preferences over experts and quacks and failures of contingent reasoning

Yan Xu

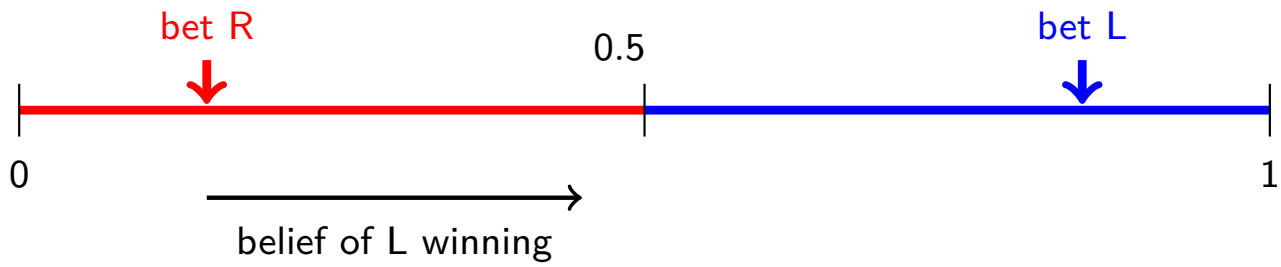
July 2020

Research question

- How do people choose and evaluate tests?
  - example: investment advisers; doctors; medical tests ...
  - timing: before receiving advice
- Quacks vs. experts: useless or useful tests
- Can people distinguish quacks and experts among competing tests?
- Are they over-paying for quacks and under-paying for experts?
- What are the mechanisms of choosing quacks?

An example of choosing tests

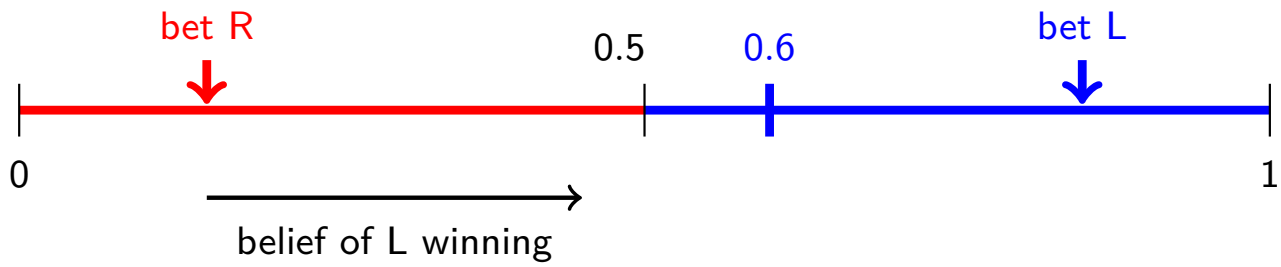
Task: bet race results b/w two horses (L and R) to win a prize π.



An example of choosing tests

Task: bet race results b/w two horses (L and R) to win a prize π.

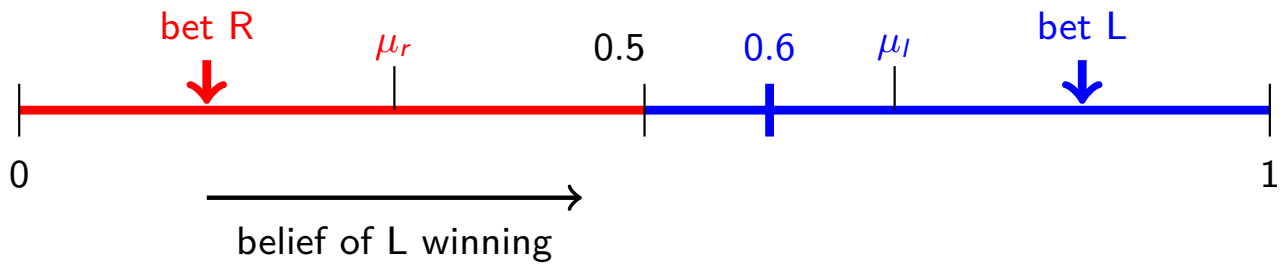
- In the past 100 races, L won 60 times and R won 40 times.



An example of choosing tests

Task: bet race results b/w two horses (L and R) to win a prize π.

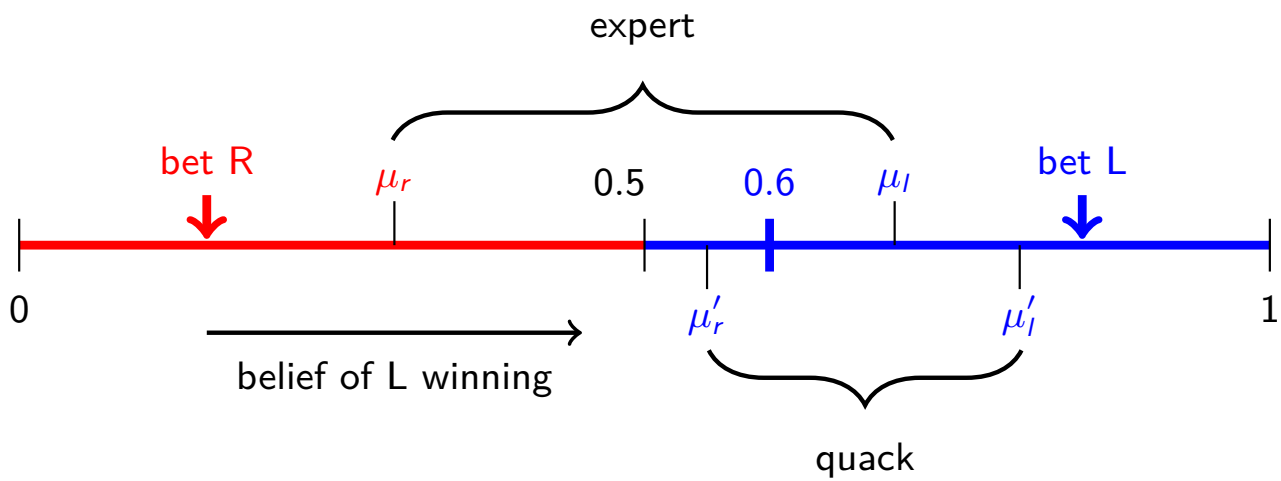
- In the past 100 races, L won 60 times and R won 40 times.
- Also consult an analyst, get an advice, and then make the bet.
- An analyst with performances of giving correct suggestions:
  - In 60 races L won, he predicted 42 times correctly (70%).
  - In 40 races R won, he predicted 18 times correctly (45%).
- How much would you like to pay for an advice from this analyst?



An example of choosing tests

Task: bet race results b/w two horses (L and R) to win a prize π.

- In the past 100 races, L won 60 times and R won 40 times.
- Also consult an analyst, get an advice, and then make the bet.
- An analyst with performances of giving correct suggestions:
  - In 60 races L won, he predicted 42 times correctly (70%).
  - In 40 races R won, he predicted 18 times correctly (45%).
- How much would you like to pay for an advice from this analyst?

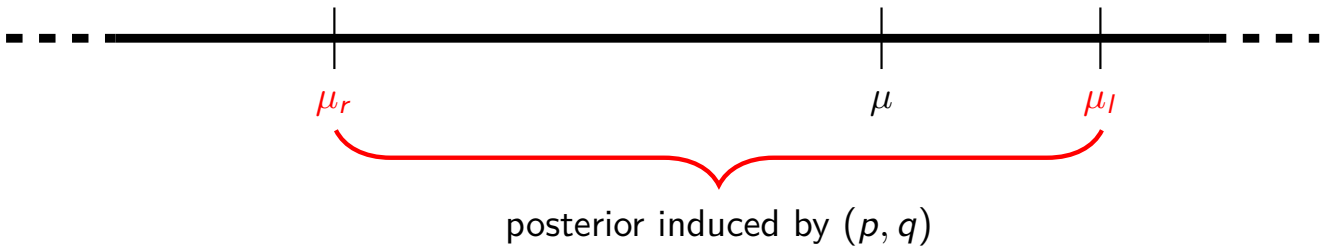


Setup: states, signals, and tests

- Two states $\omega \in \{L,\ R\}$ and two signals $s \in \{l, r\}$
- The prior $\mu \equiv \mathbb{P}(\omega = L)$
- The action space is binary: $u(a, \omega) = \pi \mathbb{I}_{a=\omega}$.
  — optimal action is to bet the state she believes $\geq 1/2$.
- 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 $\mu_r \equiv \mu(\omega = L \mid s = r)$

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

- Throughout: $\mu \geq 1/2$.
- Admissible tests: $p \geq 1 - q \iff \mu_r^{Bayes} \leq \mu \leq \mu_l^{Bayes}$



- 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 DM's ex-ante winning probability of π is:

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

Mechanisms

A DM fails to distinguish quacks and experts because she:

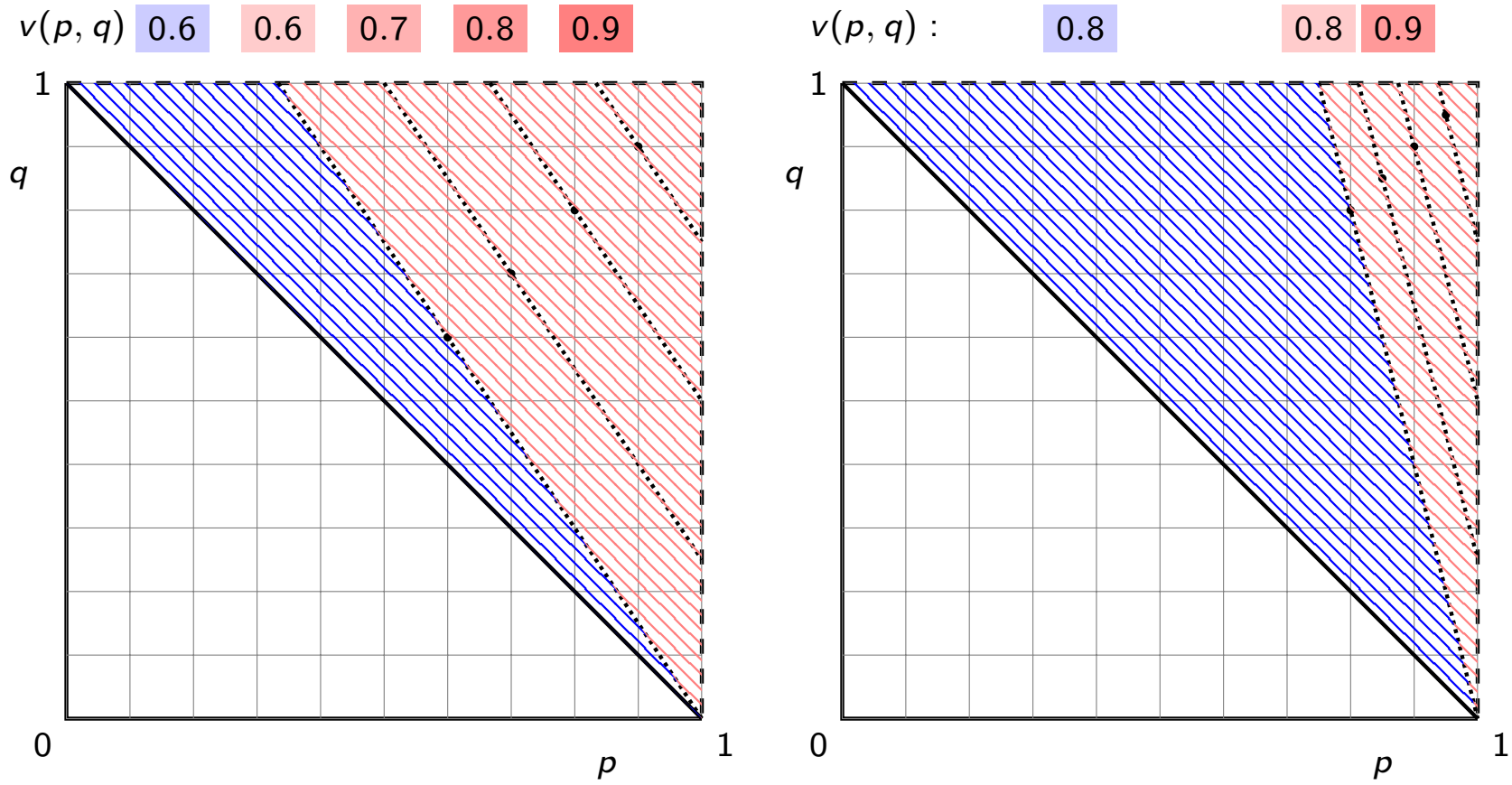
1. fails to **update beliefs** as a Bayesian
2. chooses **sub-optimal actions** given her beliefs
3. has intrinsic **preference** over certain types of tests
4. lacks **contingent reasoning** in the interaction b/w test and action

Intuition for contingent reasoning: A signal that changes the optimal action should be more valuable. A test is useful in providing an opportunity to contingent actions.

- quack: induced posteriors do not change the optimal action (pooling): $a^*(l) = a^*(r)$
- expert: induced posteriors do change the optimal action (separating): $a^*(l) \neq a^*(r)$

This paper: elicits preferences over tests and identify channels

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

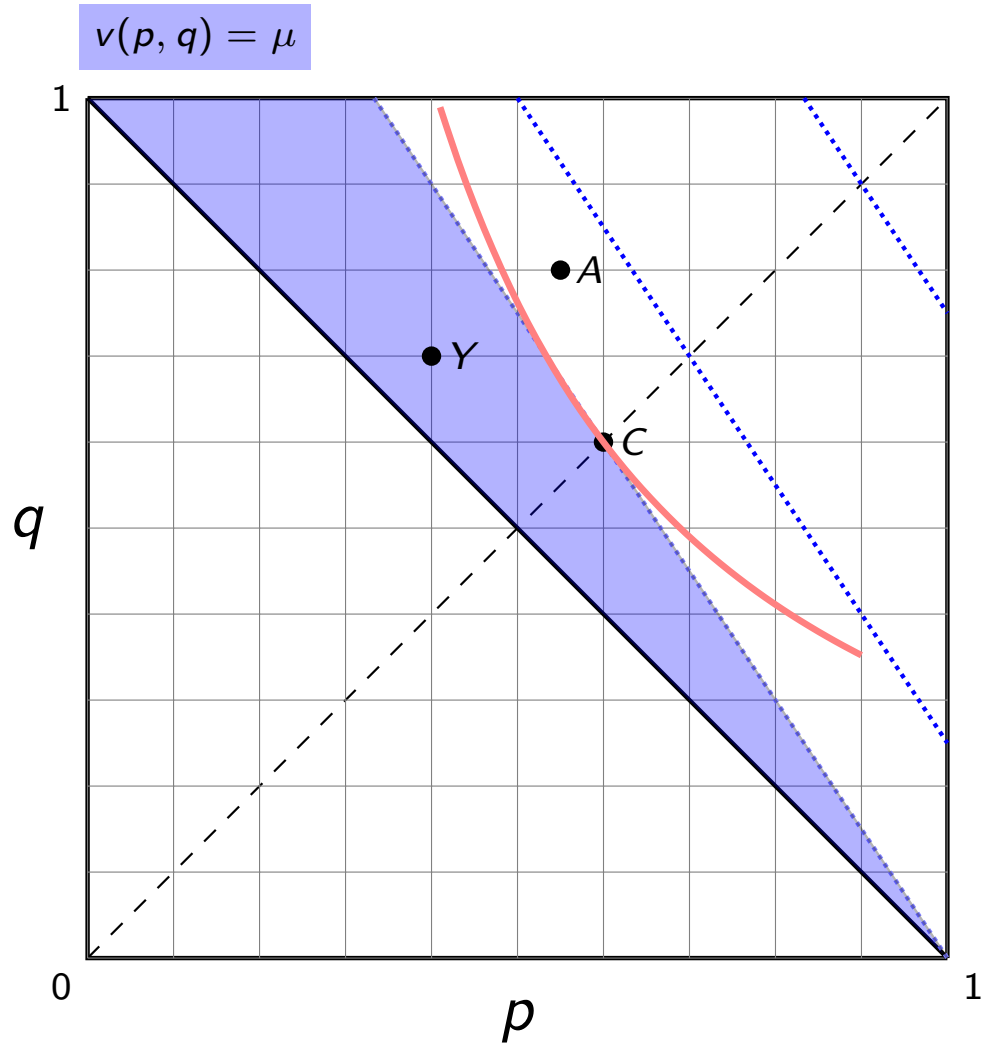


(a) Prior $\mu = 0.6$ (b) Prior $\mu = 0.8$

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

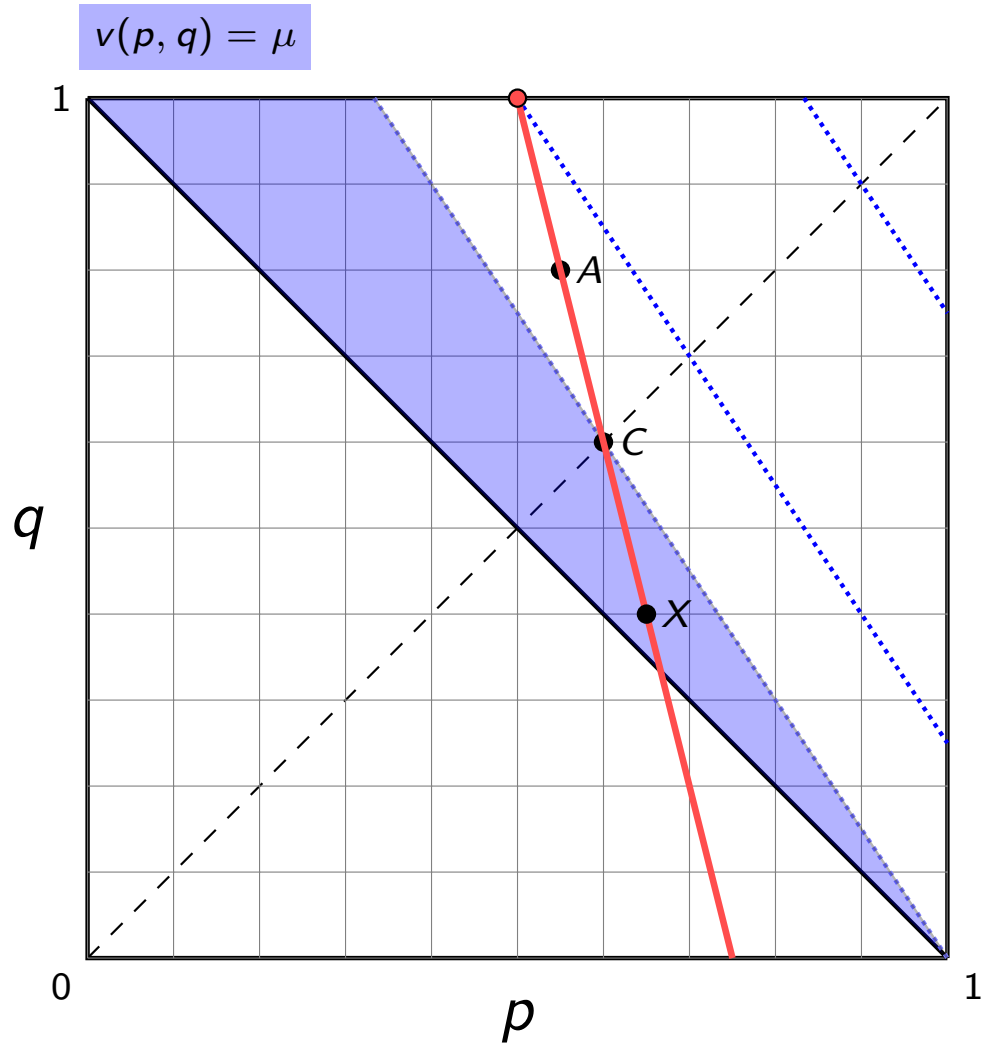
$v(p, q; \mu)$: expected winning probability of the prize under prior μ and test (p, q)

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



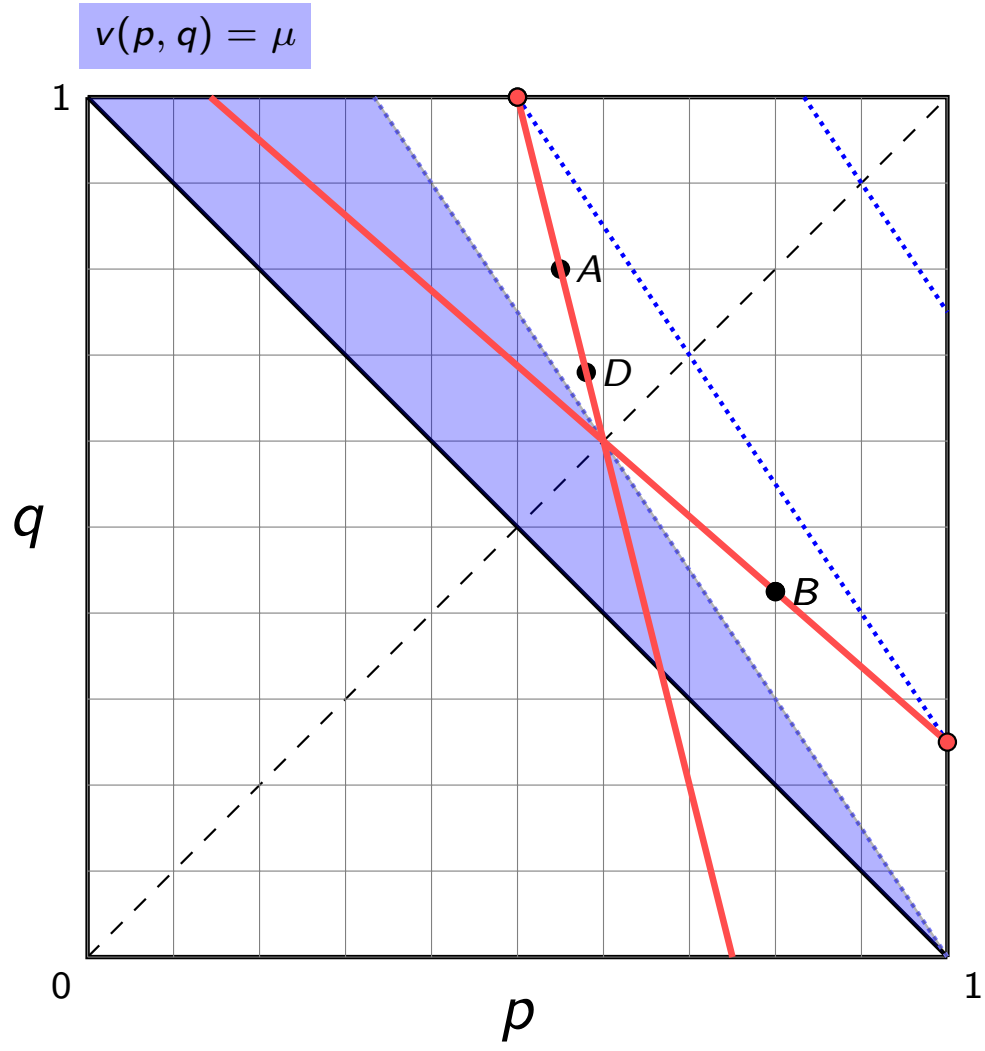
- Alternative interpretation: trade off Type I and Type II errors: $1 - p$ vs. $1 - q$
- The receiver operating characteristic (ROC) curve: p vs. $1 - q$

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



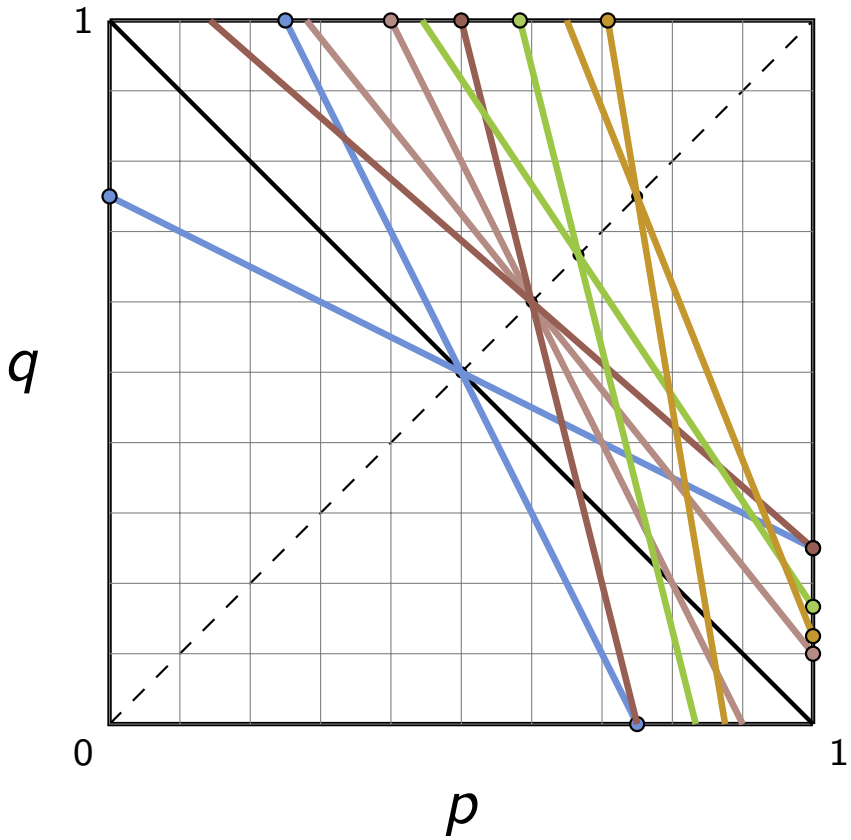
- Alternative interpretation: trade off Type I and Type II errors: $1 - p$ vs. $1 - q$
- The receiver operating characteristic (ROC) curve: p vs. $1 - q$

Eliciting preference over tests: paired linear budgets

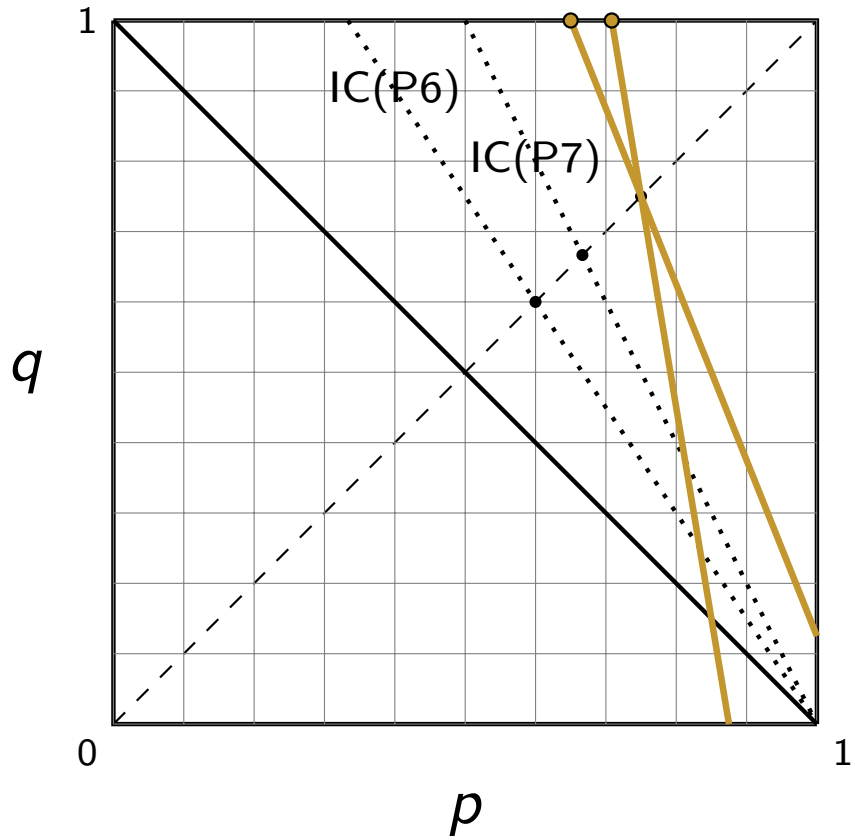


- A and B are equally useful expert tests
- If choose (D, B), the DM has intrinsic preference for p-skewed tests

Budgets for 14 rounds of tasks



(a) Budgets for P1-P5



(b) Budgets for P6-P7

Experimental task: bet horse L or horse R

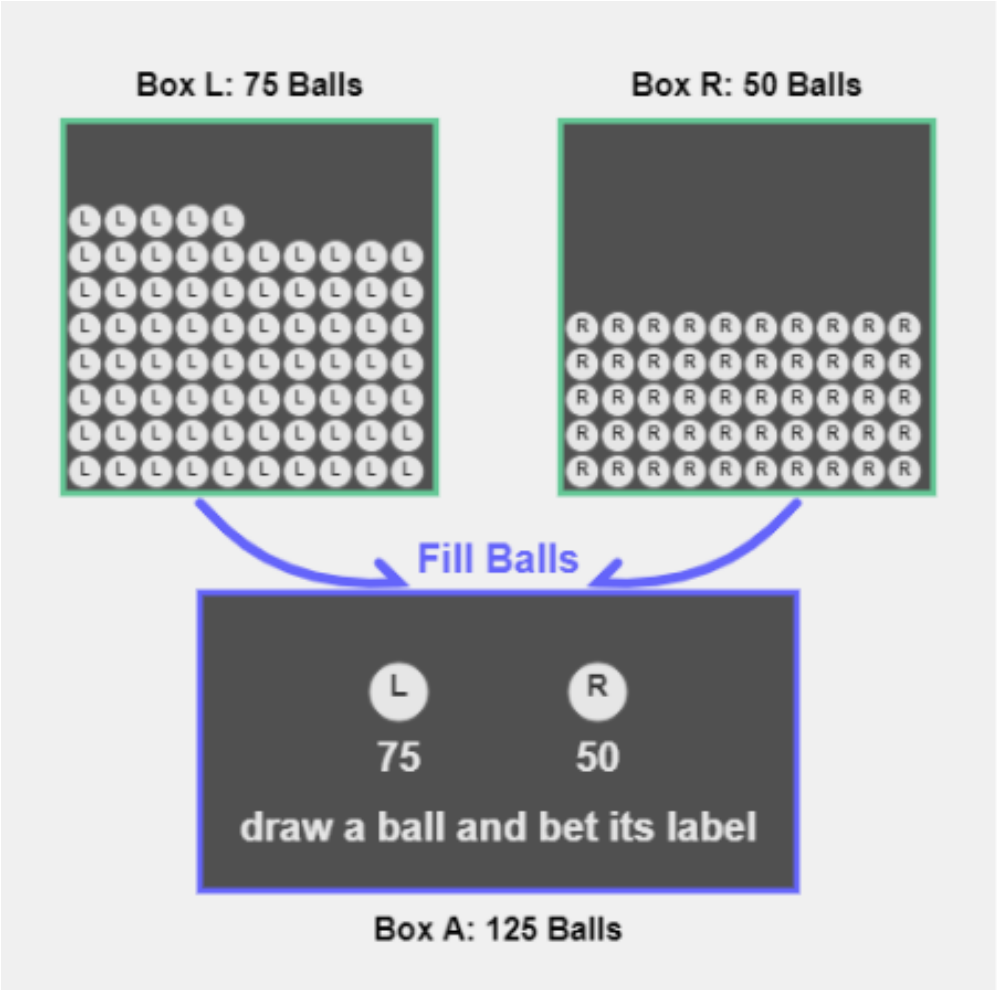
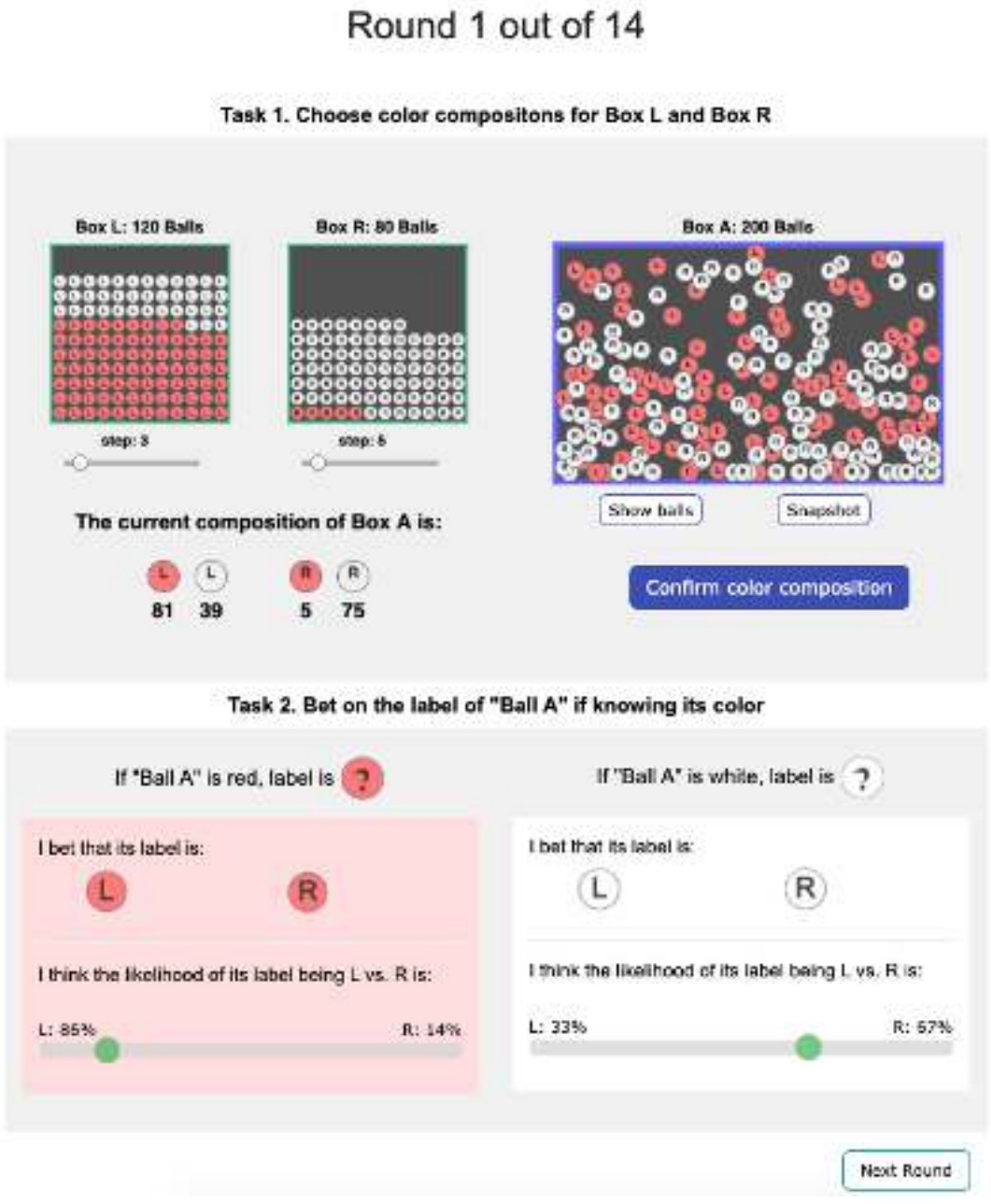


Figure: One ball (called "Ball A") will be drawn from Box A. The task is to bet its label to be either L or R. Correct bet wins a prize of £10; otherwise the payoff is 0.

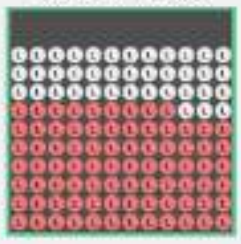
Experimental task: choose a test on a budget through a coloring task

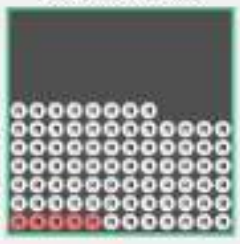


Random pay one out of fourteen rounds

Your payment

The random round is 1. Here are your choices in this round.





"Ball A" has been drawn from Box A: R

The mathematician thinks the likelihood of its label being L vs. R is:

L: 34% R: 66%

The current composition of Box A is:

L L R R

81 39 5 75

If "Ball A" is white, I bet that its label is: R

I think the likelihood of its label being L vs. R is:

L: 33% R: 67%

Your total Payment is: £15.50

= £4.00 for showing up + £10.00 for your bet choice + £1.50 for your likelihood estimation.

Please share us thoughts about how you make the color and the bet choices:

Confirm

Identifying different channels and experimental procedures

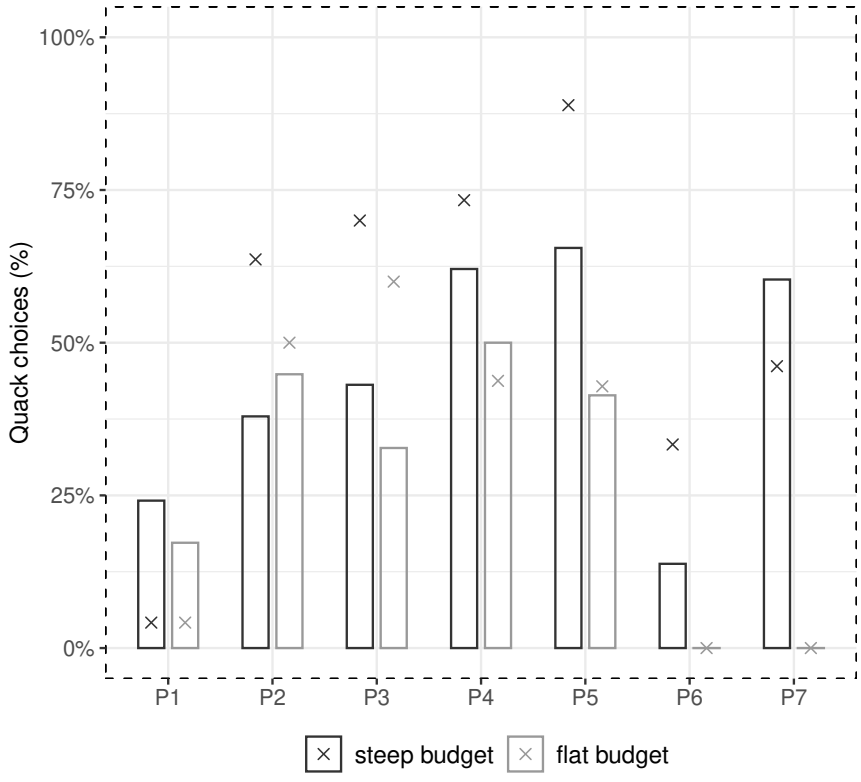
Identifications:

- 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

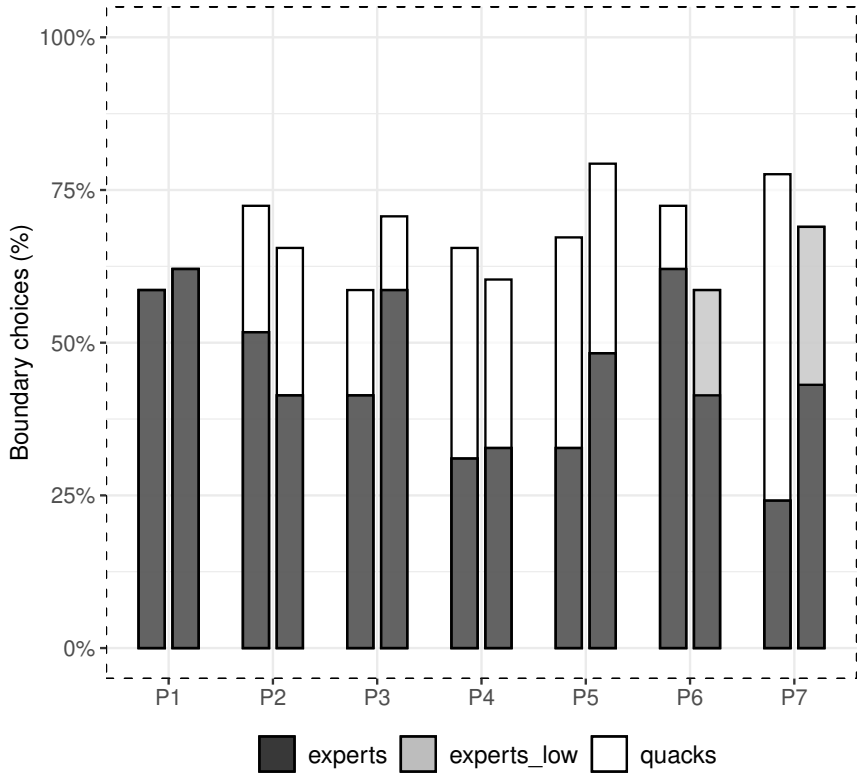
Procedures:

- 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

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

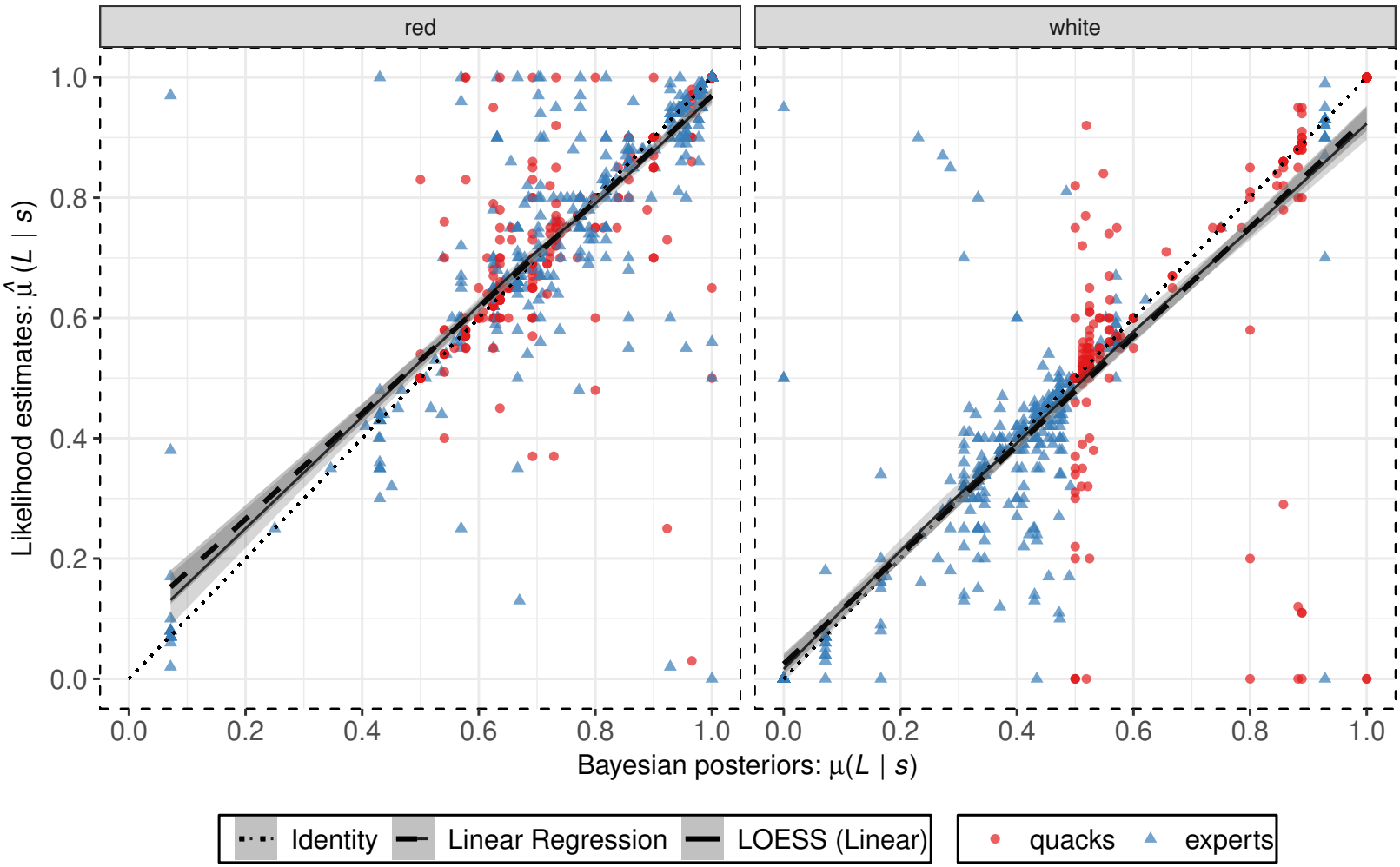
What are the consequences of choosing quacks and non-optimal experts?

	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

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

- Result: posteriors are close to Bayesian ones: 93% earn a bonus



- Result: 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: subjects are choosing the optimal bets.
- Result: best-responding biases cannot explain quack choices.

Alternative definitions of expert and quack tests

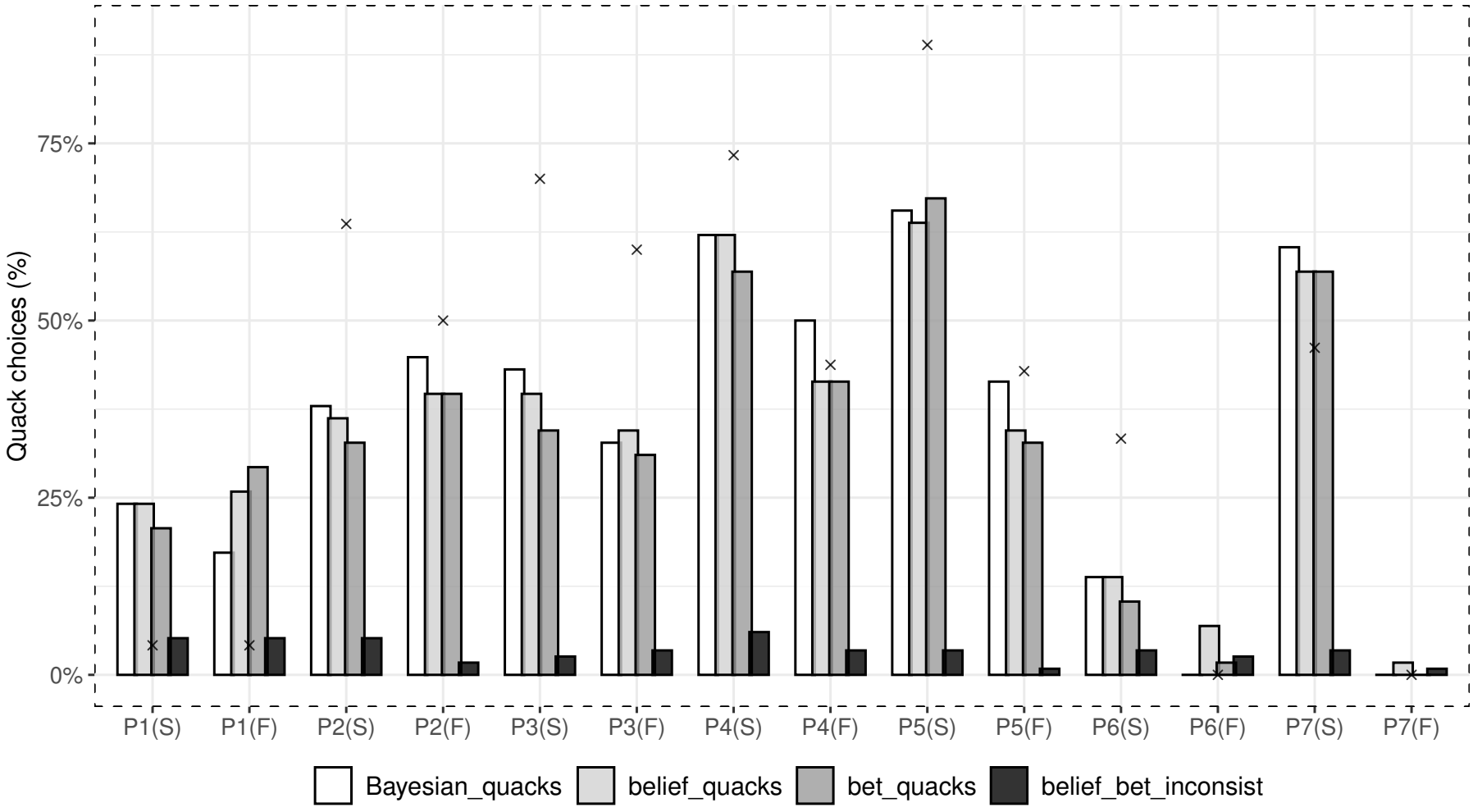


Figure: The histogram of quack choices under alternative definitions.

Channel 3: are quack choices explained by intrinsic preferences?

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

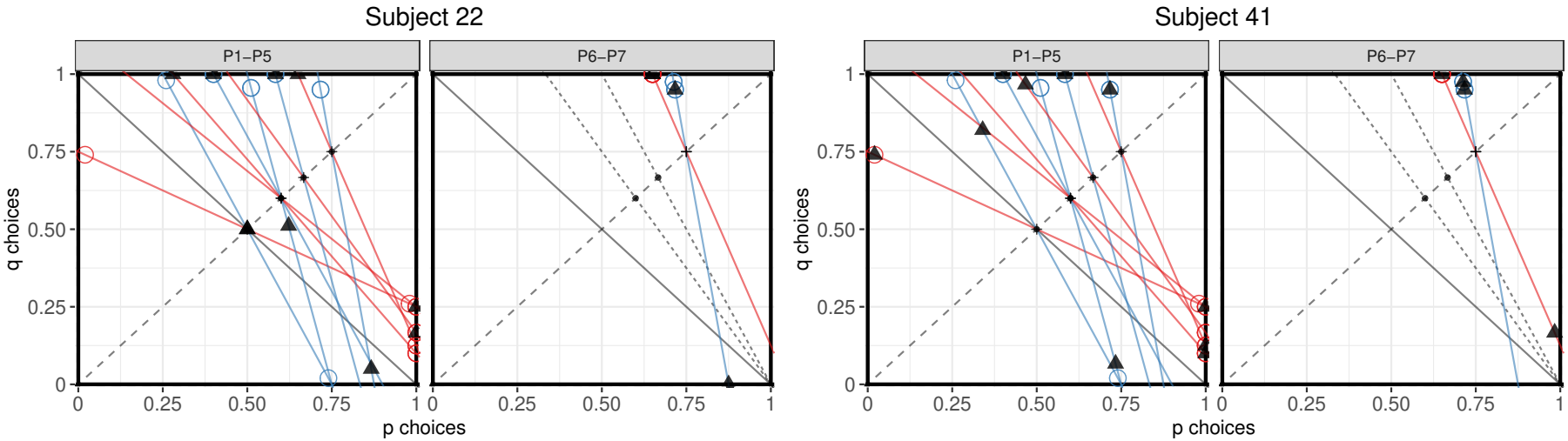
$\Rightarrow$ construct attributes measures and examine their distributions/predictability

- 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: quacks choices cannot be justified by intrinsic preferences

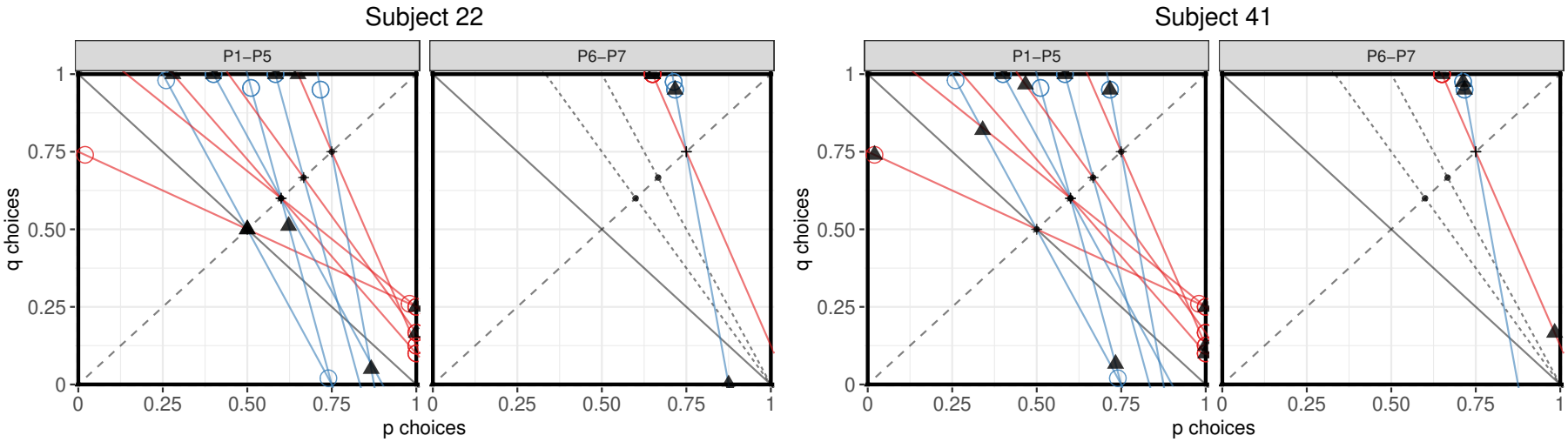
Channel 4: are quack choices explained by the lack of contingent reasoning?

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



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

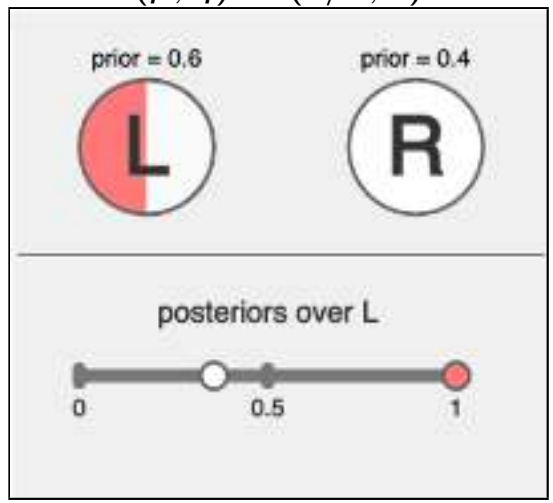
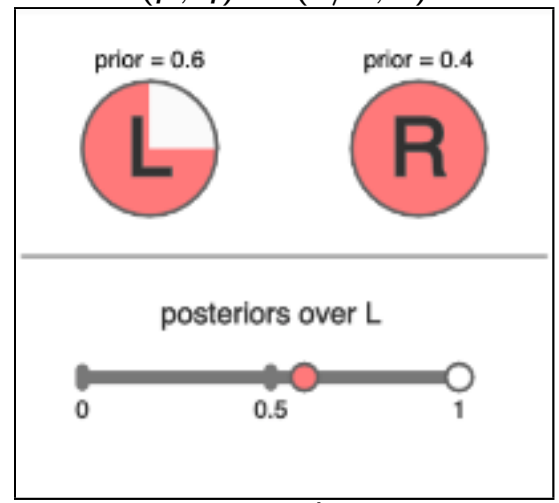
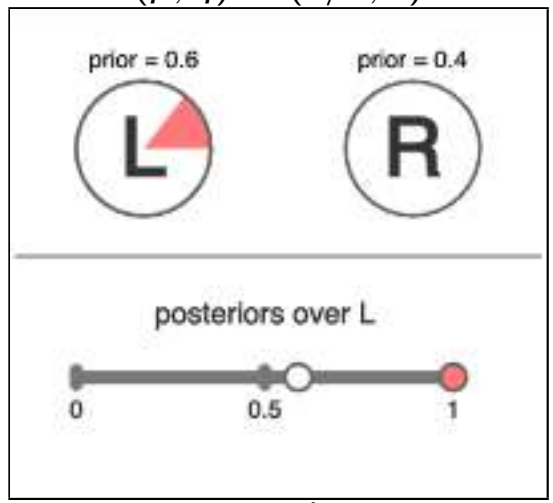
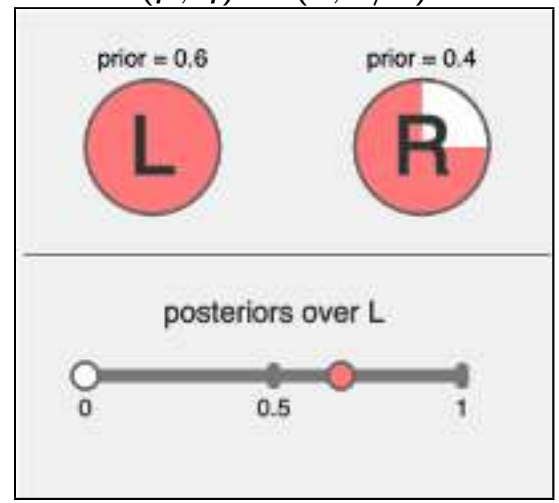
Figure: "I tried to somewhat increase the difference between two boxes"



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"

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

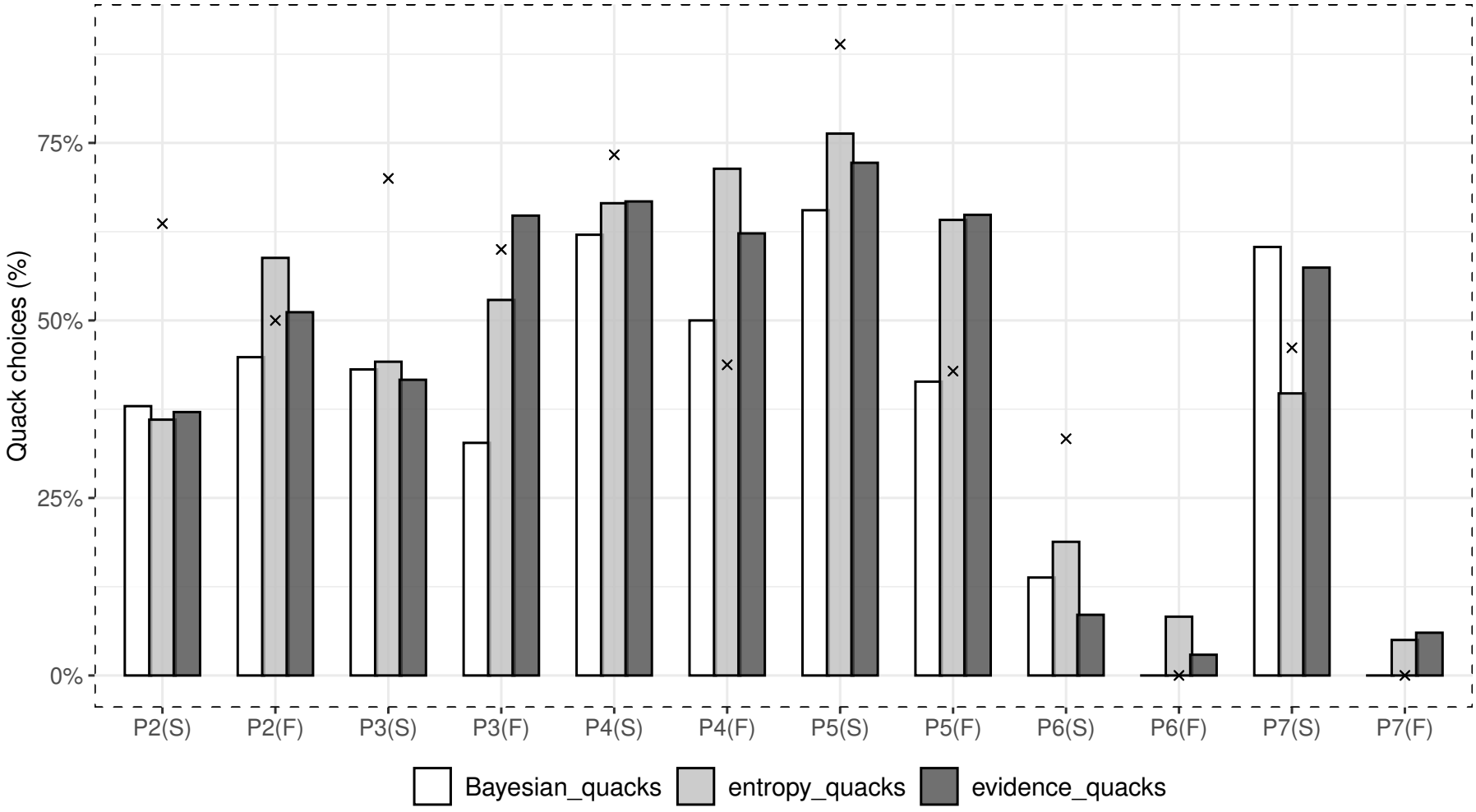
	Top-left border	Bottom-right border
Steep budget	$(p, q) = (1/2, 1)$	$(p, q) = (3/4, 0)$
q cheaper	 expert	 quack
Flat budget	$(p, q) = (1/7, 1)$	$(p, q) = (1, 1/4)$
p cheaper	 quack	 expert

Predict 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·	-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·	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

Predict the quack choice rate for each decision rule

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



- three decision rules explain the choice of border tests
- quack choices are by-products of simple decision rules
- $\Rightarrow$ failure of contingent reasoning

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, new experiment, rich choice set, 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.

Conclusion

- people fail to distinguish experts and quacks on a large scale
- quack choices are not driven by updating bias, sub-optimal actions, and intrinsic preferences
- they use entropy-reducing and evidence-separating decision rules $\Rightarrow$ border tests $\Rightarrow$ over-paying for quacks but accurately paying for experts
- people generally lack the contingent reasoning in information processing

Extensions and implications:

- general setup: asymmetric prize; non-binary signals and states;
- supply side of tests: market of ex-ante information
- strategic interactions: communication, persuasion, contract theory
- de-biasing: standard methods have a bound, new methods on reasoning?

Thanks for you patience!