Revealed Preferences over Experts and Quacks and Failures of Contingent Reasoning

Yan Xu (yan.xu@univie.ac.at)

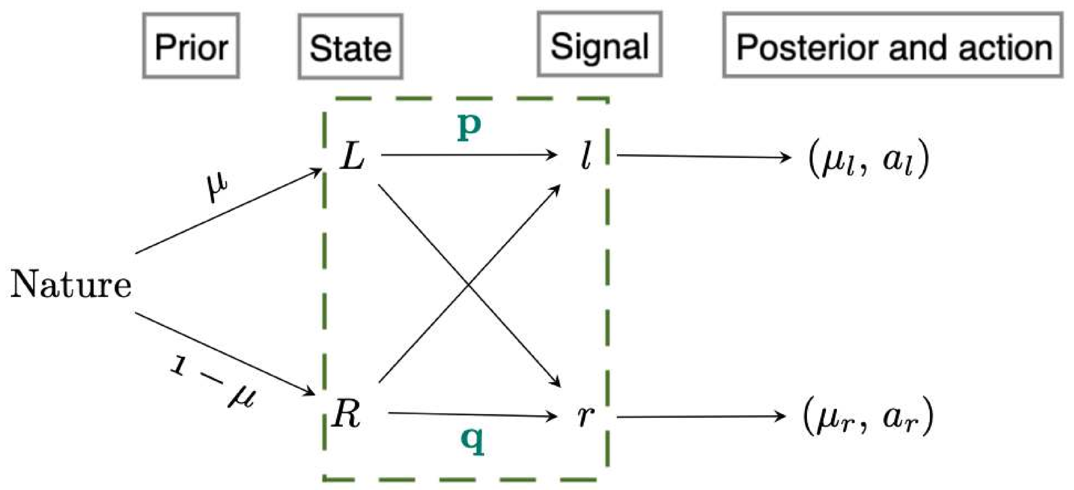
University of Vienna

Motivation: a radiologist's problem of choosing a test

Imagine a radiologist aims to figure out whether a patient has a tumor or not. She can send the patient to take a medical test, e.g., CT, MRI, X-ray, etc. Which test should she choose?

- **Decision setup:** There are two states of the world $\omega \in \{L, R\}$. Nature determines the state with an objective probability μ. Assume the radiologist's prior of tumor is $\mu(\omega = L)$.
- **Characterizing tests:** Each test generates either a positive or a negative diagnosis result (binary signal $s \in \{l, r\}$). We characterize tests by pairs (p, q), wherein $p \equiv \mathbb{P}(s = l \mid \omega = L)$ is the true positive rate, and $q \equiv \mathbb{P}(s = r \mid \omega = R)$ is the true negative rate.
- **Decision timing:** The radiologist chooses a test **before** knowing the diagnosis result. When signal s realizes, the radiologist updates her belief μ_s about tumor and determines the optimal treatment a_s. We consider binary action $a_s \in \{L, R\}$.
- **Payoffs:** The radiologist wins a prize π if the chosen treatment matches the state. Therefore, she will bet the state of which her posterior is higher than 50%. Assume the radiologist wants to maximize the chance of winning the prize.
- **Similar scenarios:** People select financial advisors, horse analysts, news sources, statistical trials, or hypothesis test methods when inferring the payoff relevant state.

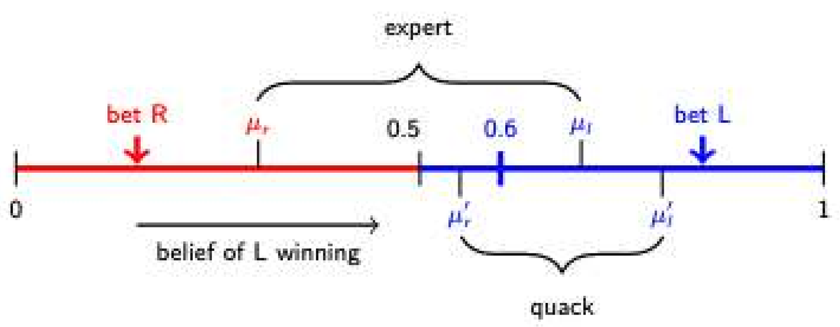
In the literature, tests are also called signal technologies, information structures, information sources, or Blackwell experiments. Given the prior, a test specifies a signal generation process.



Choice set of tests

Expert vs. Quack: how does a test improve the decision problem?

Each test induces two posteriors, which further support the optimal action for each signal and determine the expected winning probability of the prize. If a test induces posteriors on the same (different) side of the action threshold 1/2, it is ex-ante useless (useful) for the decision problem.



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

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

A DM fails to distinguish between quacks and experts because she may:

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

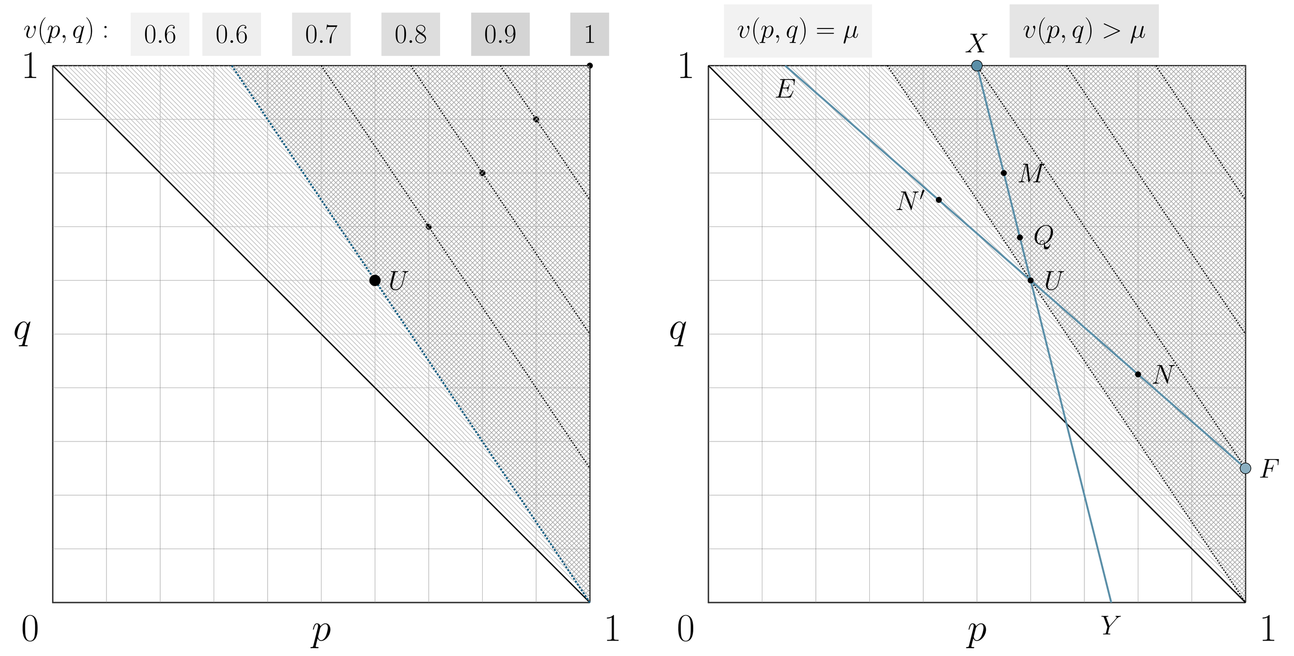
Intuition of the contingent reasoning in choosing tests: a test is (ex-ante) useful because it provides an opportunity to contingent actions on different signal realizations.

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

Experimental design: eliciting test preferences via linear budgets

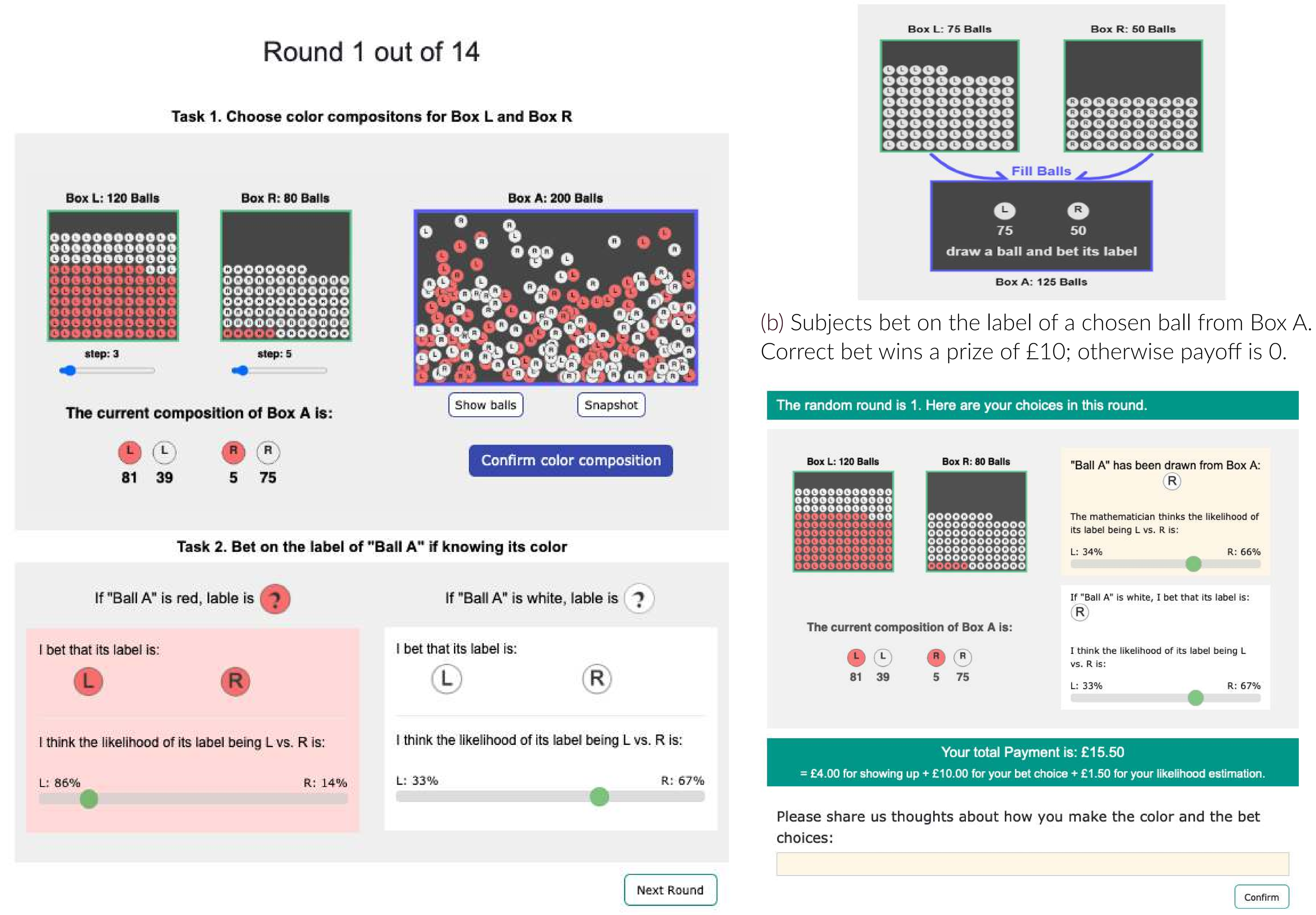
Throughout, we assume $\mu \geq 1/2$ and consider tests $\mathcal{T} \equiv \{(p, q) \mid p + q \geq 1, 0 \leq p, q \leq 1\}$.

- A rational Bayesian DM's indifference curves over tests are parallel lines. The equation $v(p, q; \mu) = \mu$ induces a line that partitions the choice set into expert and quack tests.
- A non-rational DM's indifference curves may be non-linear but should be downward-sloping, reflecting a trade-off between p and q or equivalently Type I and Type II error.
- Therefore, people's preferences over tests will be revealed through their choices of bundle (p, q) on budgets $p + mq = z$. Each 1% increase in q implies a decrease of m% in p.
- Moreover, I use paired budgets to identify and measure intrinsic preference. For instance, test M and N are equally useful but skew to different state-specific accuracies.



(a) Indiff. curves and partition of expert and quack tests (b) Test choices on paired linear budgets EF and XY

Experimental interface: each color composition of two boxes is a test



(a) Before pooling all balls, subjects color balls of two boxes through sliders. Choosing a color composition is equivalent to choosing a test on a budget.

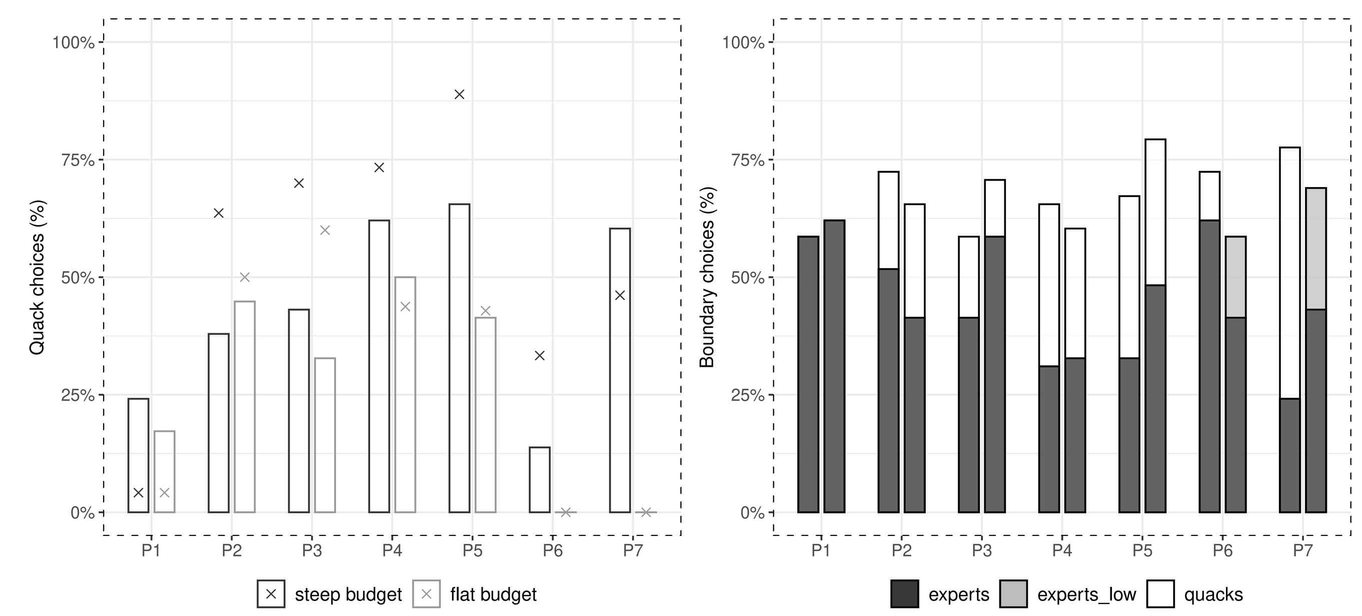
(b) Subjects bet on the label of a chosen ball from Box A. Correct bet wins a prize of £10; otherwise payoff is 0.

(c) Payment screen. Random pay one out of fourteen rounds. Each round corresponds to a linear budget.

- **State and signal** are the label and the color of the chosen ball.
- **Budget trade-offs** are achieved by linking two sliders such that every three more red balls in Box L ($p \uparrow$) brings five more red balls in Box R ($q \downarrow$).
- **Auxiliary tasks** of likelihood estimates and bet choices elicit subjects' posteriors and optimal action after each signal. They are also part of the incentive scheme.
- Recruit 64 students on Prolific. Average payoff is £11.25, and average duration is 45 minutes.

Result: failure to distinguish between quack and expert tests

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



(c) Frequency of quack choices (d) Frequency of border choices

Mechanisms: why do people choose quack tests?

1. Are quack choices explained by **belief updating biases**?
   - Result 1.1: subjects' reported posteriors are very close to Bayesian ones
   - Result 1.2: small updating biases cannot explain quack choices
2. Are quack choices explained by **sub-optimal actions**?
   - Result 2.1: subjects chose actions that best-respond to their beliefs
   - Result 2.2: small best-responding biases cannot explain quack choices
3. Are quack choices explained by **intrinsic preferences**?
   - construct attribute measures for $skew(p, q)$ and examine their distributions or predictability
   - Result 3.1: subjects exhibited intrinsic preferences for tests
   - Result 3.2: but quacks choices cannot be justified by intrinsic preferences

Subjects reported three popular decision rules on how they chose color 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 diff. in red and white b/w 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"

The three rules focus on either test characteristics, the induced posteriors, or the unconditional probability of signals. None of them considers how tests provide contingency value by influencing the optimal actions for different signals. They explain the choice of border tests, though.

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, favoring border tests.
- People in general lack the contingent reasoning in choosing and evaluating tests

Related work on information structures and contingent reasoning

1. When there are strategic interactions between an information designer and a DM ...
   - Commitment and communication in Bayesian persuasion: theory and experiment (draft coming soon)
   - Surprisingly, subjects recognize the contingency value of tests when persuading others or being persuaded.
2. When a DM designs her information environment to maintain her motivated beliefs ...
   - Motivated self-persuasion (work in progress)