Belief identification with state-dependent utilities

Elias Tsakas

Maastricht University

LOFT
July 2022

Background and contribution

- An agent has actual subjective probabilistic beliefs
  - How likely is it that the defendant is guilty?
  - How likely is it that you are caught without a train ticket?
  - How likely is it that you are better than median driver?
- Can we identify these beliefs?
- Fundamental identification problem (well-known among theorists since Savage and Aumann): This is impossible using traditional choice data when the agent has stakes on the underlying event.
- Two types of solutions within the literature:
  - Practically-oriented: pretend the agent has no stakes on the event
  - Theoretically-sound: go beyond traditional choice data (complex)
- In this paper, we will propose a novel (portable) method that circumvents the identification problem in a simple way.

Decision theoretic framework

- Two states: $\Theta = \{\theta_0, \theta_1\}$
- Acts: $f : \Theta \to X$
- (Actual) belief (probability of θ_1): μ
- Preferences over acts: $\succeq$

$\left.\begin{array}{l} \mu \\ \succeq \end{array}\right\}$ Primitives

- Assume there is a (state-dependent) SEU representation involving the actual beliefs:

there is $u = (u_0, u_1)$ such that $\mathbb{E}_\mu(u(\cdot))$ represents $\succeq$.

What is the identification problem?

- There are also SEU representations involving other beliefs:
  - Take any other $\tilde{\mu}$, and define $\tilde{u} = (\tilde{u}_0, \tilde{u}_1)$ by

$$\tilde{u}_0 = \frac{1-\mu}{1-\tilde{\mu}} u_0 + c \text{ and } \tilde{u}_1 = \frac{\mu}{\tilde{\mu}} u_1 - \frac{\tilde{\mu}}{1-\tilde{\mu}} c$$

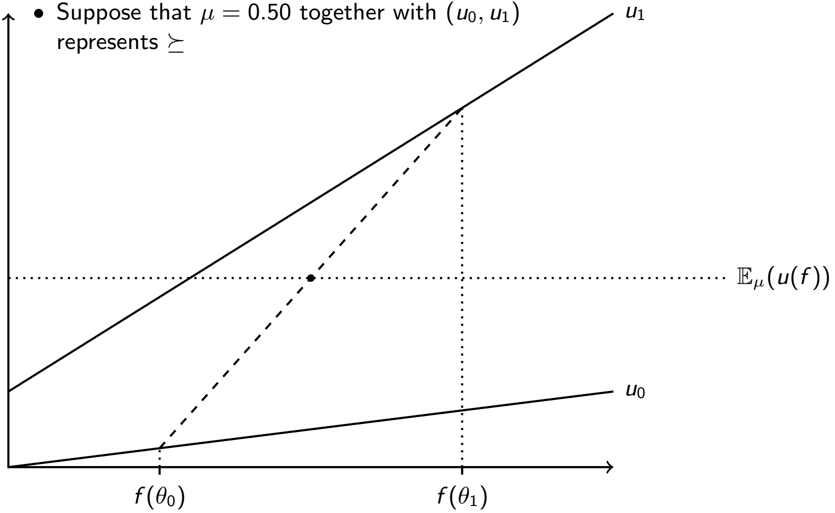
  - Then, for every act f,

$$\mathbb{E}_{\mu}(u(f)) = \mathbb{E}_{\tilde{\mu}}(\tilde{u}(f))$$

  - So, $\mathbb{E}_{\tilde{\mu}}(\tilde{u}(\cdot))$ also represents $\succeq$
- We cannot identify beliefs from choices over acts

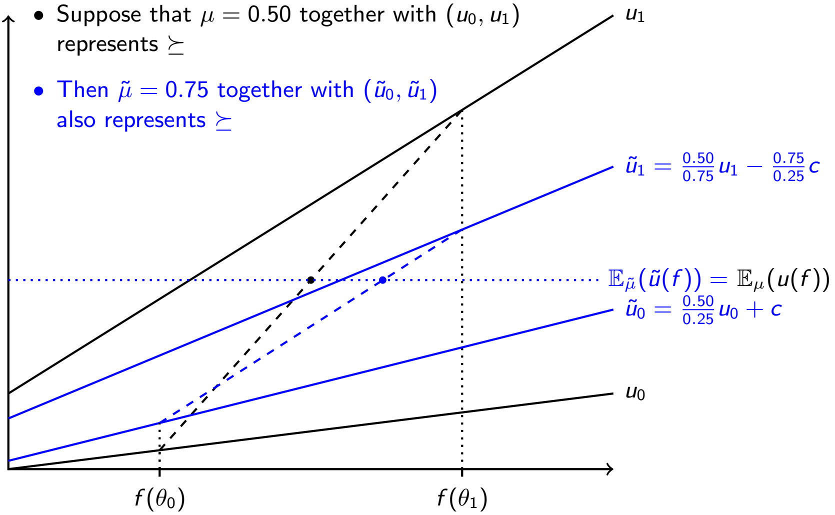
The identification problem graphically

- Suppose that $\mu = 0.50$ together with (u_0, u_1) represents $\succeq$



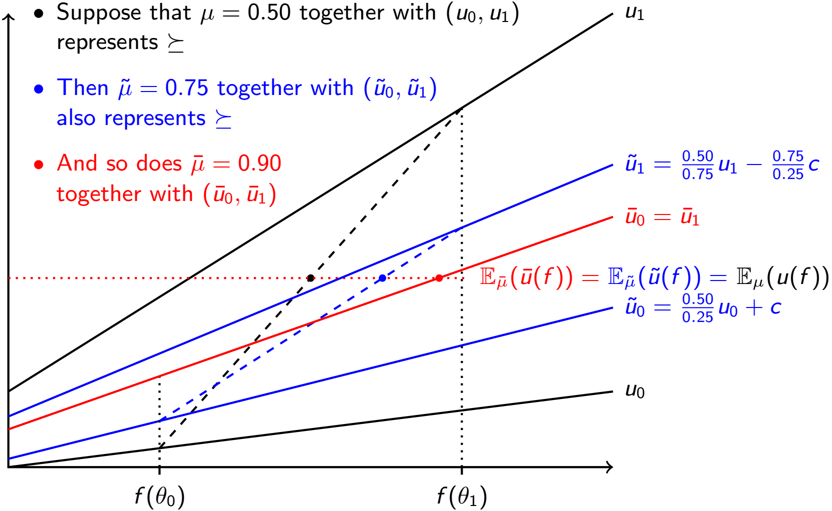
The identification problem graphically

- Suppose that $\mu = 0.50$ together with (u_0, u_1) represents $\succeq$
- Then $\tilde{\mu} = 0.75$ together with $(\tilde{u}_0, \tilde{u}_1)$ also represents $\succeq$



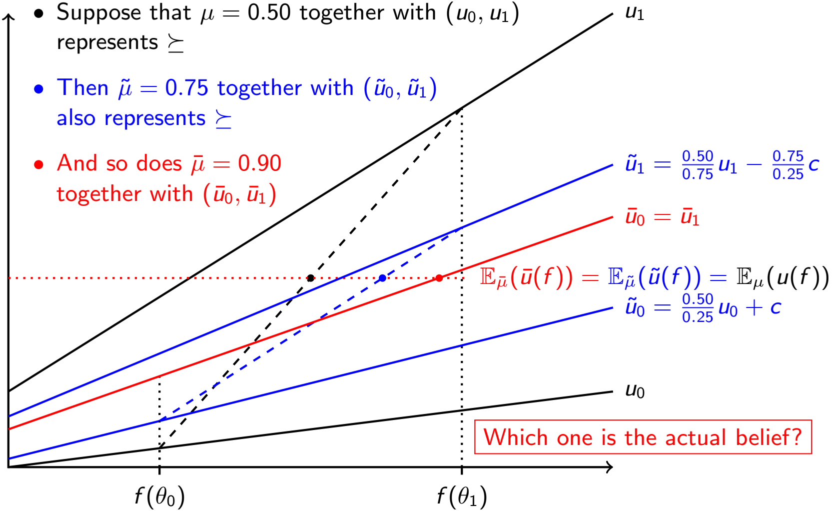
The identification problem graphically

- Suppose that $\mu = 0.50$ together with (u_0, u_1) represents $\succeq$
- Then $\tilde{\mu} = 0.75$ together with $(\tilde{u}_0, \tilde{u}_1)$ also represents $\succeq$
- And so does $\bar{\mu} = 0.90$ together with $(\bar{u}_0, \bar{u}_1)$



The identification problem graphically

- Suppose that $\mu = 0.50$ together with (u_0, u_1) represents $\succeq$
- Then $\tilde{\mu} = 0.75$ together with $(\tilde{u}_0, \tilde{u}_1)$ also represents $\succeq$
- And so does $\bar{\mu} = 0.90$ together with $(\bar{u}_0, \bar{u}_1)$



The early (practically-oriented) solution

- Assume state-independent utilities:
  1. If a state-independent SEU representation exists, assume that $\bar{\mu}$ is the actual belief
  2. Use some standard incentive-compatible elicitation mechanism (e.g., binarized scoring rule) to elicit $\bar{\mu}$
- There are multiple problems with this solution:
  - A state-independent SEU representation may not even exist
  - Even if it exists, it is only reasonable to assume that $\bar{\mu}$ is the actual belief if the agent does not have stakes in the underlying event
- Savage recognized that this was an issue: *"the problem is serious, but I am willing to live with it until something better comes along"*
  - In all fairness, if we only care about providing foundations to subjective probability or about predicting choices among acts, this is not a serious problem.

Subsequent (theoretically-sound) solutions

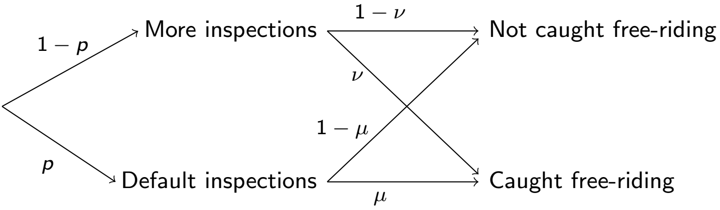
- Go beyond traditional choice data, e.g.,
  - Dréze (1961): the analyst can influence the state realization
  - Fishburn (1973); Karni (1992,1993): choices conditional on different events
  - Karni, Schmeidler & Vind (1983): choices given hypothetical beliefs
  - Schervish, Seidenfeld & Kadane (1990): the agent compares lotteries at different states
  - Lu (2019): the agent updates beliefs using information that the analyst provides
- Implementation is very complex (sometimes practically impossible)
- No consensus on one of these solutions
- The problem is very difficult, and still open!!!

Our approach

1. We enlarge the state space along a second dimension S, so that
   (a) Marginal distribution p over S is commonly known
   (b) S correlated with Θ
   (c) Given each θ, the agent does not have stakes on S
2. We elicit conditional beliefs about S given each θ.
   - Truthful elicitation is possible, because of (c)
3. We can identify the beliefs on Θ
   - Identification is possible, because of (a) and (b)

Example 1: Stochastic intervention

- $\Theta = \{$Caught free-riding, Not caught free-riding$\}$
- $S = \{$More inspections, Default inspections$\}$

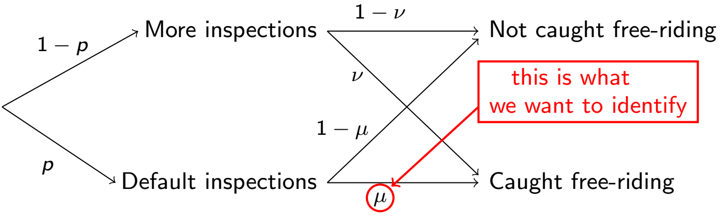


Theorem (Identification result)

where $p_0 = P(Default \mid Not\ caught)$ and $p_1 = P(Default \mid Caught)$ is what we elicit, and $p = P(Default)$ is publicly announced.

Example 1: Stochastic intervention

- $\Theta = \{\text{Caught free-riding, Not caught free-riding}\}$
- $S = \{\text{More inspections, Default inspections}\}$

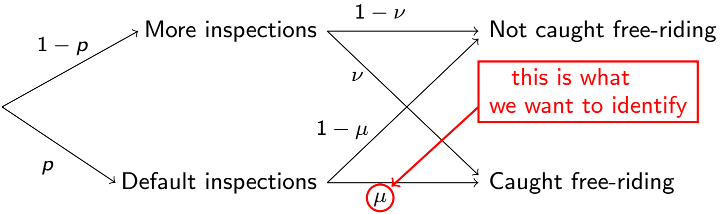


Theorem (Identification result)

where $p_0 = P(Default \mid Not\ caught)$ and $p_1 = P(Default \mid Caught)$ is what we elicit, and $p = P(Default)$ is publicly announced.

Example 1: Stochastic intervention

- $\Theta = \{$Caught free-riding, Not caught free-riding$\}$
- $S = \{$More inspections, Default inspections$\}$



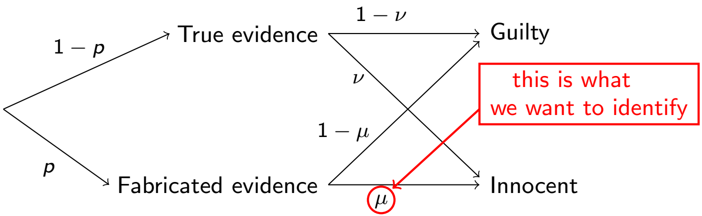
Theorem (Identification result)

$$\mu = \frac{p_1(p-p_0)}{p(p_1-p_0)} \quad \textit{and} \quad \nu = \frac{(1-p_1)(p-p_0)}{(1-p)(p_1-p_0)}$$

where $p_0 = P(Default \mid Not\ caught)$ and $p_1 = P(Default \mid Caught)$ is what we elicit, and $p = P(Default)$ is publicly announced.

Example 2: Stochastic truth of evidence

- $\Theta = \{\text{Guilty, Innocent}\}$
- $S = \{\text{True evidence, Fabricated evidence}\}$



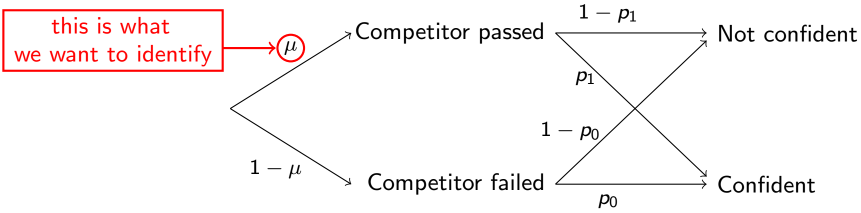
Theorem (Identification result)

$$\mu = \frac{p_1(p - p_0)}{p(p_1 - p_0)} \quad \text{and} \quad \nu = \frac{(1 - p_1)(p - p_0)}{(1 - p)(p_1 - p_0)}$$

where $p_0 = P(\text{Fabricated} \mid \text{Guilty})$ and $p_1 = P(\text{Fabricated} \mid \text{Innocent})$ is what we elicit, and $p = P(\text{Fabricated})$ is publicly announced.

Example 3: Partition of the population

- $\Theta = \{\text{Competitor passed, Competitor failed}\}$
- $S = \{\text{Confident, Not confident}\}$



Theorem (Identification result)

$$\mu = \frac{p - p_0}{p_1 - p_0}$$

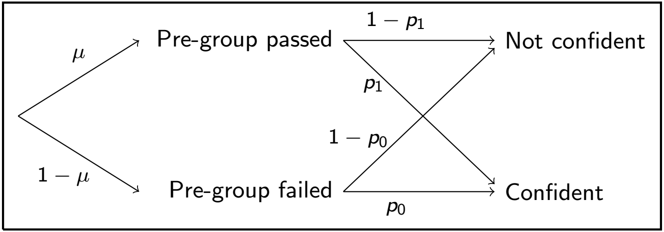
where $p_0 = P(\textit{Confident} \mid \textit{Failed})$ and $p_1 = P(\textit{Confident} \mid \textit{Passed})$ is what we elicit, and $p = P(\textit{Confident})$ is commonly known.

Is this approach appealing?

- Theoretically, yes: It solves a long-standing problem in a simple way!!!
- Empirically: it seems so!!!
  1. Flexibility in which second dimension to use? Yes!!
  2. Can it be adjusted without Bayesian updating? Yes!!
  3. Do we restrict elicitation mechanism? No!!

Proof of concept: Pilot experiment

- Pre-group: answer to Math Question (Θ) and stated Confidence (S).
- Main group: informed total probability of "Confident" is $p = 55\%$, and report **direct belief** (μ) and conditional beliefs (p_0 and p_1).
- **Indirect belief** is $((p - p_0)/(p_1 - p_0))$ computed



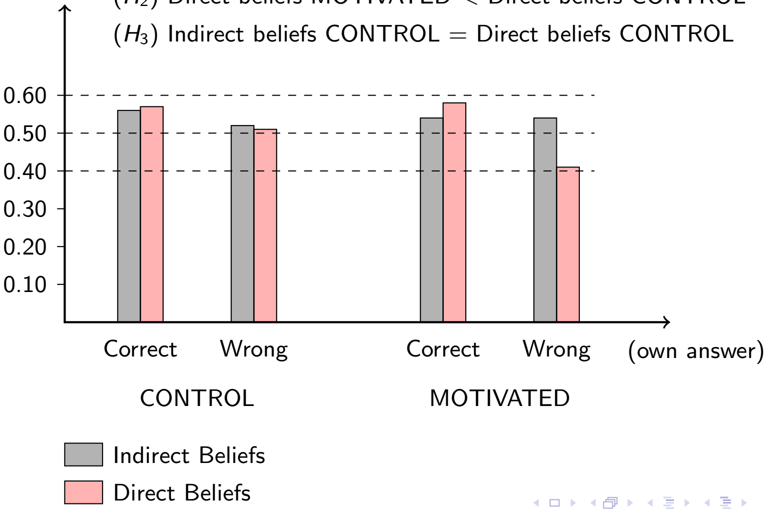
- Two treatments (CONTROL and MOTIVATED) differ in knowledge of their own result, to increase saliency of comparision.
  - (H_1) Indirect beliefs MOTIVATED = Indirect beliefs CONTROL
  - (H_2) Direct beliefs MOTIVATED < Direct beliefs CONTROL
  - (H_3) Indirect beliefs CONTROL = Direct beliefs CONTROL

Good news: it seems to work

(H_1) Indirect beliefs MOTIVATED $=$ Indirect beliefs CONTROL

(H_2) Direct beliefs MOTIVATED $<$ Direct beliefs CONTROL

(H_3) Indirect beliefs CONTROL $=$ Direct beliefs CONTROL

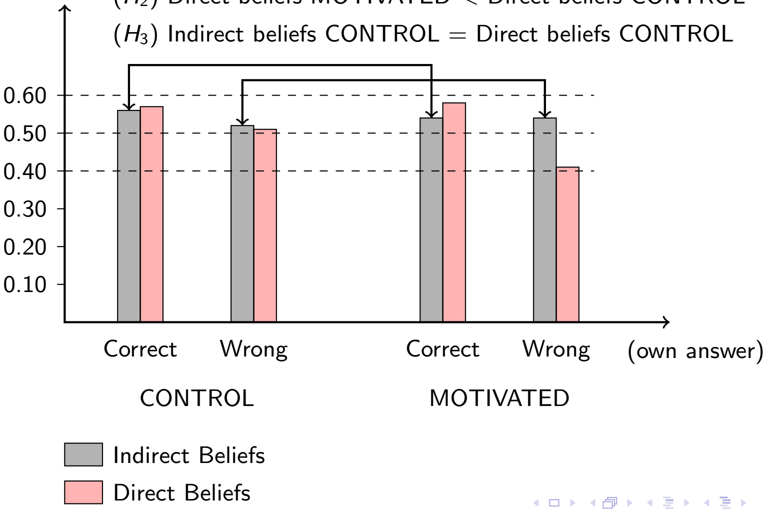


Good news: it seems to work

(H_1) Indirect beliefs MOTIVATED = Indirect beliefs CONTROL

(H_2) Direct beliefs MOTIVATED < Direct beliefs CONTROL

(H_3) Indirect beliefs CONTROL = Direct beliefs CONTROL

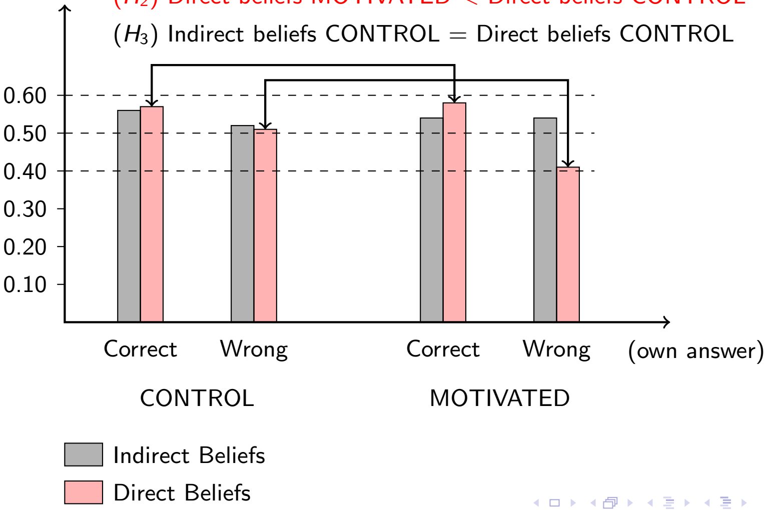


Good news: it seems to work

(H_1) Indirect beliefs MOTIVATED $=$ Indirect beliefs CONTROL

(H_2) Direct beliefs MOTIVATED $<$ Direct beliefs CONTROL

(H_3) Indirect beliefs CONTROL $=$ Direct beliefs CONTROL

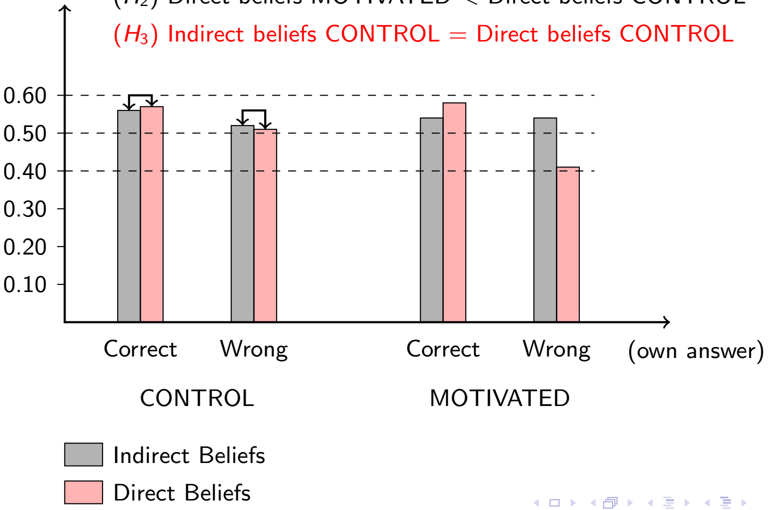


Good news: it seems to work

(H_1) Indirect beliefs MOTIVATED $=$ Indirect beliefs CONTROL

(H_2) Direct beliefs MOTIVATED $<$ Direct beliefs CONTROL

(H_3) Indirect beliefs CONTROL $=$ Direct beliefs CONTROL



Thanks for listening!!!