A robust measure of complexity

Egor Bronnikov & Elias Tsakas

Maastricht University

Berlin Behavioral Economics Seminar
May 2025

Roadmap

Roadmap

Background

- Most Economists agree that (task) complexity is a key determinant of human behavior.
- It has the potential to explain mistakes that people systematically make
- At the same time, there is no consensus on what complexity is.
- Often, it is defined in a casual way

How is complexity formalized?

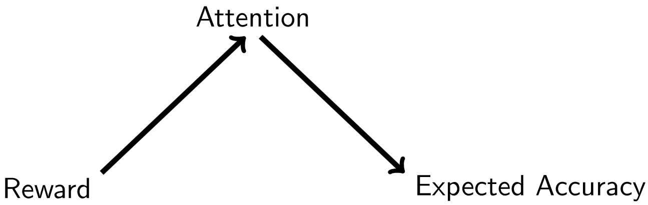
1. **Direct approach**: start with a definition, e.g.,
   - Characteristics of lotteries (Huck & Weizsäcker, 1999; Fudenberg & Puri, 2023; Enke & Shubatt, 2023; Hu, 2023; de Clippel et al., 2025)
   - Degree of contingent reasoning in mechanisms (Nagel & Saitto, 2025)
   - How pronounced tradeoffs are (Shubatt & Yang, 2025)
   - Through productivity of thinking about a task (Gabaix & Graeber, 2024)
   - Signal-to-noise ratio (Goncalves, 2024)
2. **Revealed complexity**: start with a measure/proxy, e.g.,
   1. **Direct metrics**:
      - WTP to avoid a task (Oprea, 2020),
      - response times (Wilcox, 1993; Goncalves, 2024),
      - biometrics (van der Wel & van Steenbergen, 2018).
   2. **Behavioral metrics**:
      - choice inconsistencies (Woodford, 2020).
   3. **Belief-based metrics**:
      - expected accuracy (Agranov, Schotter & Trevino, 2025; Enke & Graeber, 2023; Enke, Graeber & Oprea, 2025; Hu, 2024; ...)

Expected accuracy as a measure of complexity

- Expected accuracy: Probability to solve task correctly.
- Basic idea: $\boxed{\text{Higher complexity} \;:=\; \text{Lower expected accuracy}}$
- Reasons to use it:
  1. It is simple and intuitive!
  2. Gaining momentum in the literature!
- On the flip side, there are two important caveats:
  1. There are no choice theoretic foundations.
     - What are we actually measuring?
  2. The induced complexity order depends on the size of the reward
     - Chances to solve task A are larger than task B, if reward is high
     - Chances to solve task A are smaller than task B, if reward is low
- Thus, the following question arises:

Is it a reasonable measure of complexity?

Expected accuracy depends on reward

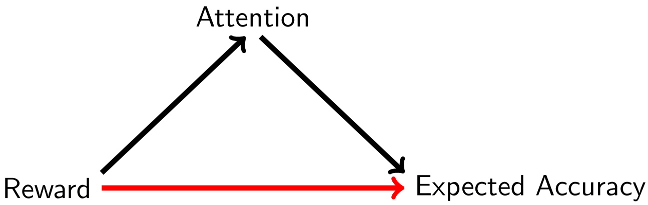


- Reward affects attentions

  (higher reward, more attention)
- Attentions affects expected accuracy

  (more attention, more likely to be correct)
- Hence, reward affects expected accuracy

Expected accuracy depends on reward



- Reward affects attentions
  (higher reward, more attention)
- Attentions affects expected accuracy
  (more attention, more likely to be correct)
- Hence, reward affects expected accuracy **non-linearly**

Roadmap

Our approach

- We take a robust approach

$$\boxed{\text{Higher complexity} \;:=\; \text{Lower expected accuracy } \textbf{for every reward}}$$

- Conservative (dominance) criterion: not all tasks will be ranked
- Complexity order:
  - (+) Intuitively appealing
  - (−) Incomplete

Preview of main conclusions

- **Theoretical conclusions**:
  1. Condition for any reasonable definition of complexity
     (based on the appeal of our criterion)
  2. Degree of uncertainty is an essential part of complexity
     (when is an exam deemed complex?)

- **Practical conclusions**:
  1. Recently popular measure of complexity is validated
     (using a lab experiment)
  2. Elicit expected accuracy for more rewards
     (using strategy method)

Our relation to the literature

1. Literature justifies belief-based measures of complexity based on common sense (Agranov, Schotter & Trevino, 2025; Enke & Graeber, 2023; Enke, Graeber & Oprea, 2025; Hu, 2024; . . .)
   - We provide theoretical foundations

2. Literature takes complexity almost as a synonym to difficulty in the corresponding context (Oprea, 2024; Nagel & Saitto, 2025; Shubatt & Yang, 2025; Gabaix & Graeber, 2024; Goncalves, 2024; . . .)
   - We identify degree of uncertainty as a novel channel

3. Literature de facto postulates that complexity is a complete order (Oprea, 2020, 2024; Nagel & Saitto, 2025; Shubatt & Yang, 2025; Gabaix & Graeber, 2024; Goncalves, 2024; Woodford, 2020; Agranov, Schotter & Trevino, 2025; Enke & Graeber, 2023; Enke, Graeber & Oprea, 2025; Hu, 2024; . . .)
   - No need to do so! We take a more conservative approach.

Roadmap

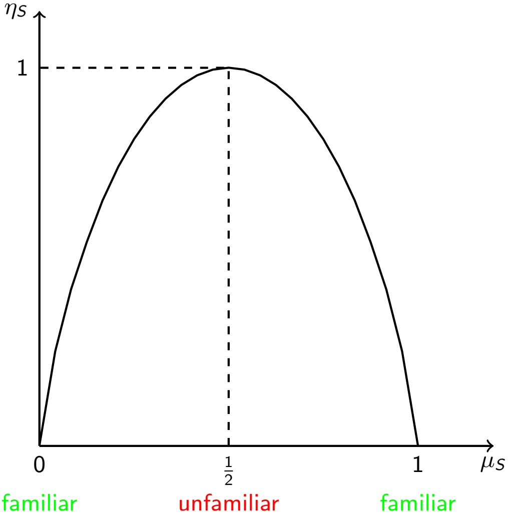
(Standard) theoretical framework

- Binary task $S = \{s_0, s_1\}$
- Set of all tasks $\mathcal{S}_0$
- Scores $Z = \{0, 1\}$
- Rewards $X = [0, \infty)$

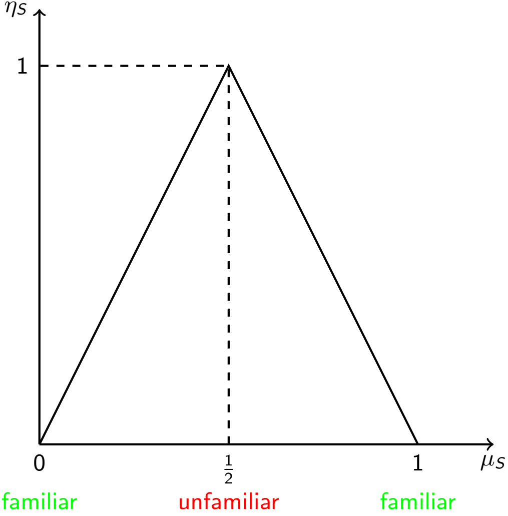
- Net utility $v_S(x) := u_S(x, 1) - u_S(0, 0) = \beta_S v(x)$
  - Task-specific subjective parameter (satisfaction)
  - Risk preferences are task-independent
  - Only for presentation purposes, fix $\beta_S := 1$ for all $S \in \mathcal{S}_0$

- Prior belief $\mu_S \in [0, 1]$ of s_1
  - Novelty of our paper to let the prior vary across tasks
- Degree of uncertainty $\eta_S = \frac{1}{\log 2} H(\mu_S)$
  - Task-specific subjective parameter (familiarity)
  - Consistent with information theory (Cover & Thomas, 2006)

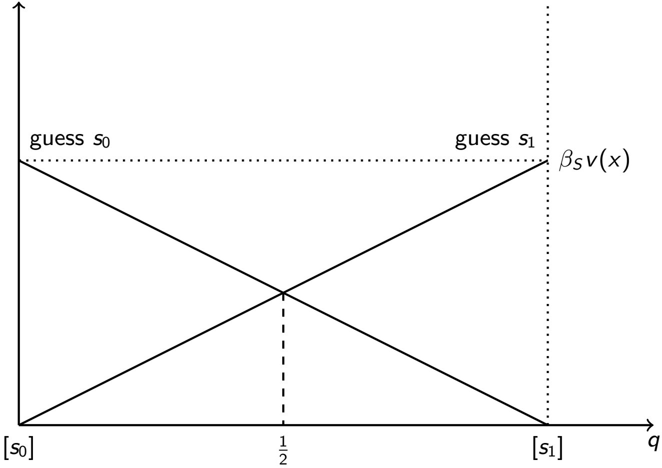
Degree of uncertainty



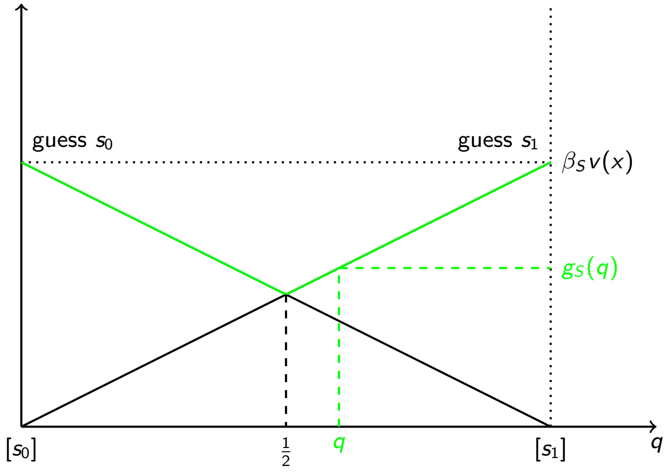
Degree of uncertainty



Utility from answering correctly

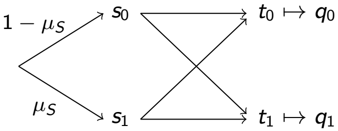


Utility from answering correctly

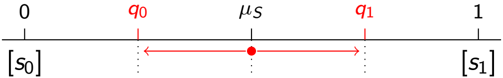


Attention

- Attention strategy: signal producing stochastic evidence

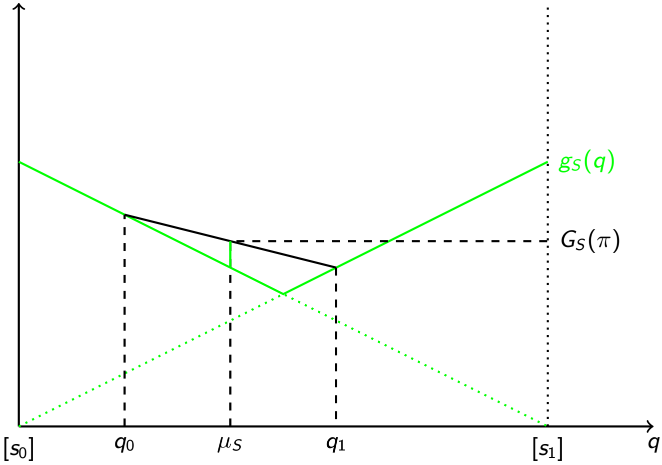


- Each attention strategy is characterized by a (mean-preserving) distribution of posteriors: $\pi \in \Delta([0,1])$ such that $\mathbb{E}_\pi(q) = \mu_S$.



- Attention has benefits and costs.
- Well-known that it is enough to focus on binary attention strategies (Matějka & McKay, 2015)

Expected benefit of attention graphically



Cost of attention

- Posterior-separable cost of attention

$$C_S(\pi) = \kappa_S \big(\mathbb{E}_\pi(c(q)) - c(\mu_S)\big)$$

    - Task-independent subjective parameter (cost of information processing)
    - Task-specific objective parameter (difficulty)
- Solid theoretical foundations (Caplin et al., 2017; Tsakas, 2020; Zhong, 2022; Denti, 2022) and support by experimental findings (Dean & Neligh, 2024)
- Symmetry of c has been axiomatized (Hébert & Woodford, 2021) and particularly natural in binary tasks

Cost of attention

- Posterior-separable cost of attention

$$C_S(\pi) = \kappa_S \big(\mathbb{E}_\pi(c(q)) - c(\mu_S)\big)$$

  - Task-independent subjective parameter (cost of information processing)
  - Task-specific objective parameter (difficulty)
- Solid theoretical foundations (Caplin et al., 2017; Tsakas, 2020; Zhong, 2022; Denti, 2022) and support by experimental findings (Dean & Neligh, 2024)
- Symmetry of c has been axiomatized (Hébert & Woodford, 2021) and particularly natural in binary tasks

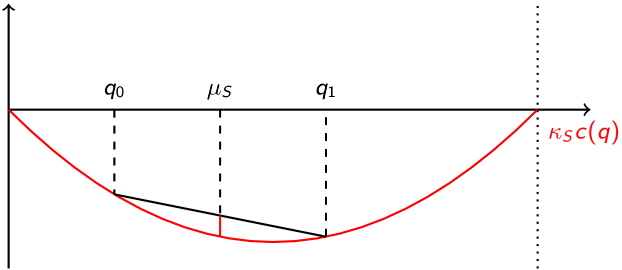
Cost of attention

- Posterior-separable cost of attention

$$C_S(\pi) = \kappa_S \left(\mathbb{E}_\pi(c(q)) - c(\mu_S) \right)$$

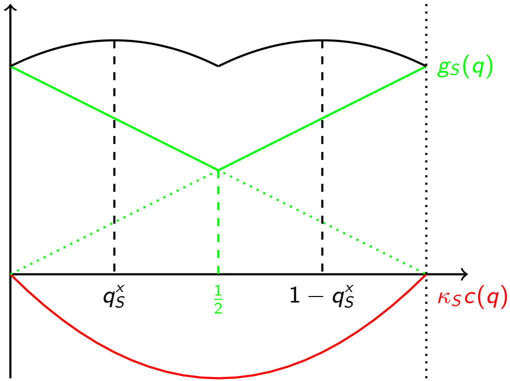
  - Task-independent subjective parameter (cost of information processing)
  - Task-specific objective parameter (difficulty)
- Solid theoretical foundations (Caplin et al., 2017; Tsakas, 2020; Zhong, 2022; Denti, 2022) and support by experimental findings (Dean & Neligh, 2024)
- Symmetry of c has been axiomatized (Hébert & Woodford, 2021) and particularly natural in binary tasks

Cost of information graphically



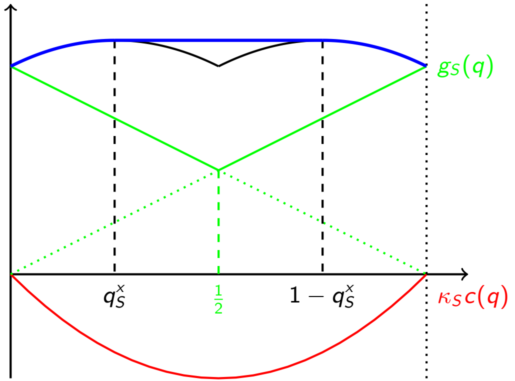
Optimal attention

- Agent's optimization problem $\max_\pi \big(G_S(\pi) - C_S(\pi)\big)$
- Solved with concavification method (Aumann & Maschler, 1995; Kamenica & Gentzkow, 2011)



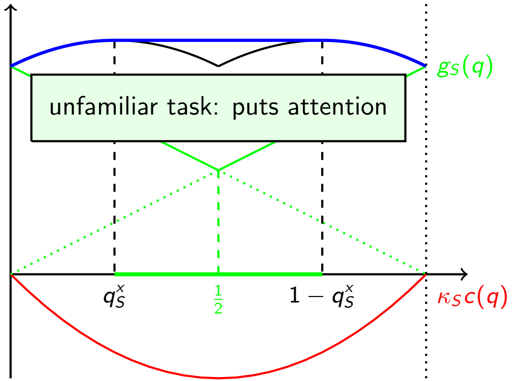
Optimal attention

- Agent's optimization problem $\max_\pi \big(G_S(\pi) - C_S(\pi)\big)$
- Solved with concavification method (Aumann & Maschler, 1995; Kamenica & Gentzkow, 2011)



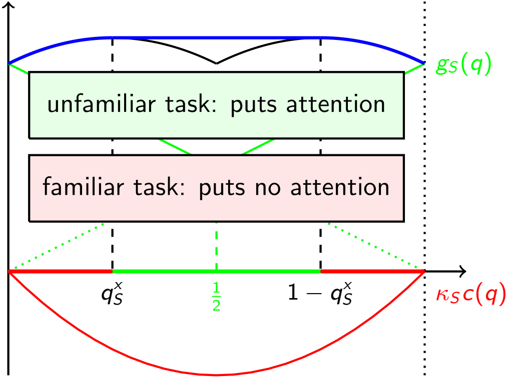
Optimal attention

- Agent's optimization problem $\max_\pi \big(G_S(\pi) - C_S(\pi)\big)$
- Solved with concavification method (Aumann & Maschler, 1995; Kamenica & Gentzkow, 2011)

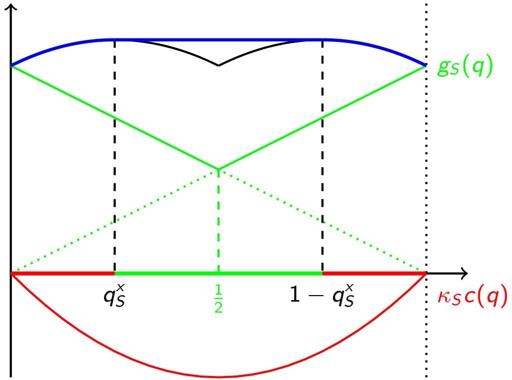


Optimal attention

- Agent's optimization problem $\max_\pi \big(G_S(\pi) - C_S(\pi)\big)$
- Solved with concavification method (Aumann & Maschler, 1995; Kamenica & Gentzkow, 2011)

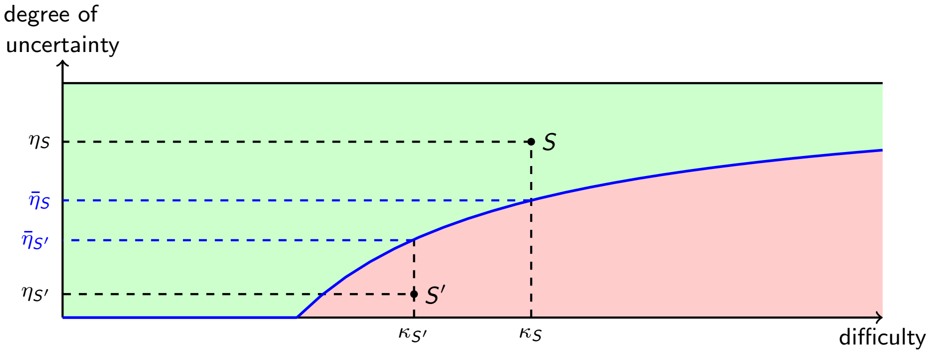


Optimal attention graphically



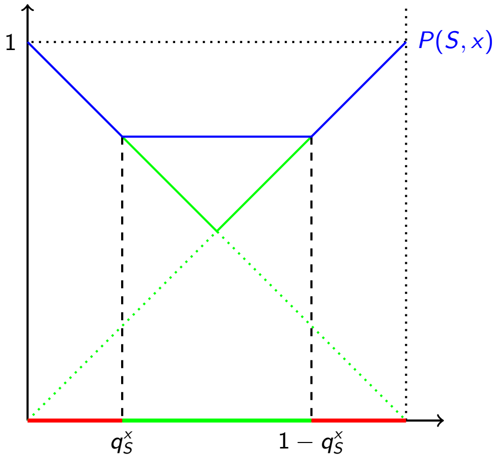
- Attention threshold q_S^x:
  - decreasing in reward (x)
  - increasing in difficulty (κ_S)
- Uncertainty threshold without reward $\bar{\eta}_S = \dfrac{H(q_S^0)}{\log 2}$

Attention map without reward



- Green area/large uncertainty ($\eta_S > \bar{\eta}_S$): attention without reward
- Red area/small uncertainty ($\eta_{S'} \leq \bar{\eta}_{S'}$): no attention without reward
- Without intrinsic incentives, the entire area is red

Expected accuracy



Robust definition of complexity

Definition

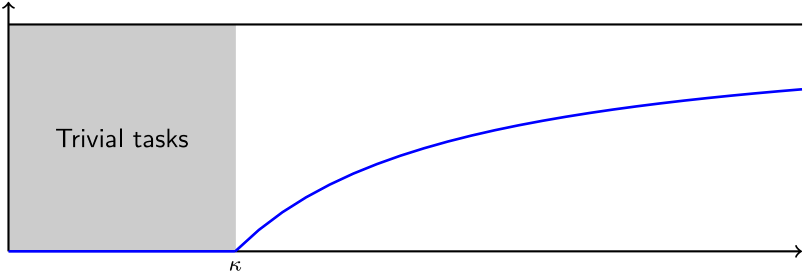
Task $S \in \mathcal{S}_0$ is more complex than $S' \in \mathcal{S}_0$ if

$$P(S, x) \leq P(S', x)$$

for all $x \geq 0$. Then, we write $S \succeq S'$.

Trivial tasks

- Task $S \in \mathcal{S}_0$ is trivial if $S' \succeq S$ for all $S' \in \mathcal{S}_0$.
- The set of non-trivial tasks is denoted by $\mathcal{S} \subseteq \mathcal{S}_0$.



- The following are equivalent:
  - S is trivial
  - $P(S, x) = 1$ for all $x \geq 0$
  - So easy that the state is learned with certainty (even without reward)

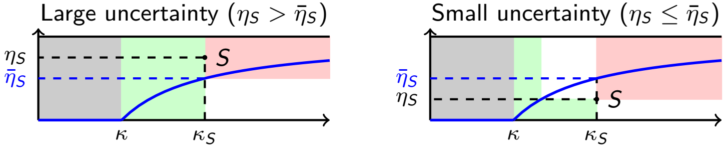
Characterization: Vector-valued representation (Ok, 2002)

Theorem (Identification)

For any pair $S, S' \in \mathcal{S}$:

$$S' \succeq S \;\Leftrightarrow\; \phi(S') \geq \phi(S),$$

where $\phi_1(S) = \kappa_S$ *and* $\phi_2(S) = \min\{\eta_S, \bar{\eta}_S\}$.



- A task is complex when it is both difficult and unfamiliar.
  (exam is complex when it is hard and and not practiced).
- Difficulty remains the primary channel: $\kappa_S > \kappa_{S'} \;\Rightarrow\; S' \not\preceq S$
  (even though not necessarily $\kappa_S > \kappa_{S'} \;\Rightarrow\; S \succeq S'$).
- Without intrinsic incentives, we have $\phi_2(S) = \eta_S$.

Characterization of incompleteness

Proposition (Single crossing)

Suppose that $S, S' \in \mathcal{S}$ are not $\succeq$-comparable, in that

- S *more difficult than* S' : $\phi_1(S) > \phi_1(S')$,
- S *more familiar than* S' : $\phi_2(S) < \phi_2(S')$.

Then, there are two thresholds $0 < x_1 < x_2 \leq \infty$ *such that:*

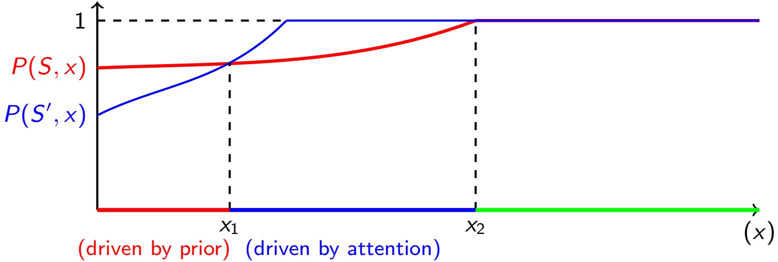
(i) $P(S,x) < P(S',x)$ *for all* $x < x_1$,
  - *For small rewards, it is more likely to solve the difficult familiar task*
  - *Not worthy paying attention, so the agent relies more on the prior*

(ii) $P(S,x) > P(S',x)$ *for all* $x_1 < x < x_2$,
  - *For large rewards, it is more likely to solve the easy unfamiliar task*
  - *Worthy paying attention, so the agent relies more on the signal*

(iii) $P(S,x) = P(S',x)$ *for all* $x \geq x_2$.

Characterization of incompleteness

- S more difficult than S' : $\phi_1(S) > \phi_1(S')$
- S more familiar than S' : $\phi_2(S) < \phi_2(S')$



- Small rewards (red area): more likely to solve the difficult familiar task
- Large rewards (blue area): more likely to solve the easy unfamiliar task
- Very large rewards (green area): both tasks solved with certainty

Detour: Eliciting the complexity order

- Can we elicit the belief $P(S, x)$?
- Binarized scoring rule pays in probability units (Hossain & Okui, 2013):

Chances to win prize	Wrong guess ($z = 0$)	Correct guess ($z = 1$)
Reported belief of z_1 (R)	$1 - \gamma R^2$	$1 - \gamma(1 - R)^2$

- Optimal report $R(S, x) \neq P(S, x)$
- Identification problem due to state-dependent utilities (Tsakas, 2025)
  - She cares about the prize y and the outcome (x, z)
  - Even if we disregard hedging opportunities
- It doesn't matter: we actually care about $\succeq$, not about $P(S, x)$:

Proposition (Elicitation)

For every pair $S, S' \in \mathcal{S}$ and every $x \geq 0$:

$$P(S, x) \geq P(S', x) \;\Leftrightarrow\; R(S, x) \geq R(S', x).$$

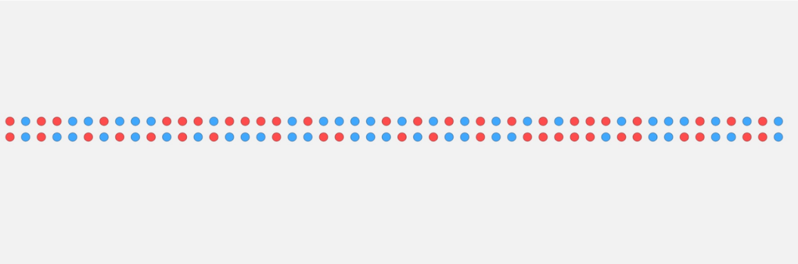
Some practical considerations

- Robustness forces us to use strategy method:
  - Elicit $R(S, x)$ for multiple $x \geq 0$
  - Elicitation must take place before the task
- There are hedging opportunities
  - Usual problem (Blanco et al., 2010)
  - It can be solved by randomly paying for one $x \geq 0$
- Alternative empirical strategy: Elicit belief about accuracy of others
  - Often simpler to implement
  - Similar idea in Bayesian markets (Baillon, 2017)
  - This is what we use in our experiment

Roadmap

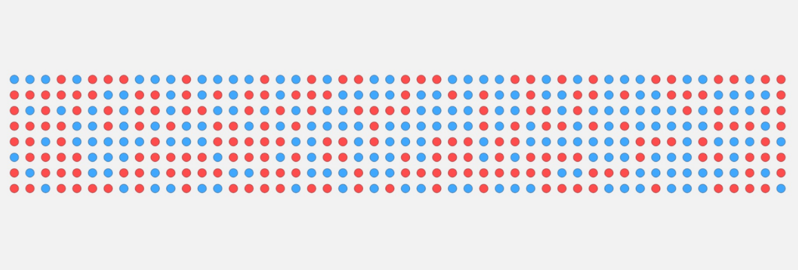
First stage

- A panel with colored balls (red and blue) is drawn
  - Easy panel (100 balls): 51 of dominant color / 49 of other color
  - Difficult panel (400 balls): 201 of dominant color / 199 of other color
- This panel is drawn from a pool (all easy or all difficult)
  - Familiar task: 8 panels of one color / 2 panels of other color
  - Unfamiliar task: 5 panels of one color / 5 panels of other color
- Participants see the drawn panel and estimate the dominant color
- Two treatments: High reward (€10) / Low reward (€0.5)



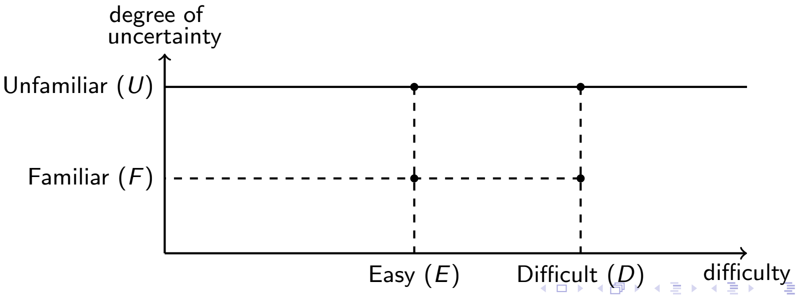
First stage

- A panel with colored balls (red and blue) is drawn
  - Easy panel (100 balls): 51 of dominant color / 49 of other color
  - Difficult panel (400 balls): 201 of dominant color / 199 of other color
- This panel is drawn from a pool (all easy or all difficult)
  - Familiar task: 8 panels of one color / 2 panels of other color
  - Unfamiliar task: 5 panels of one color / 5 panels of other color
- Participants see the drawn panel and estimate the dominant color
- Two treatments: High reward (€10) / Low reward (€0.5)



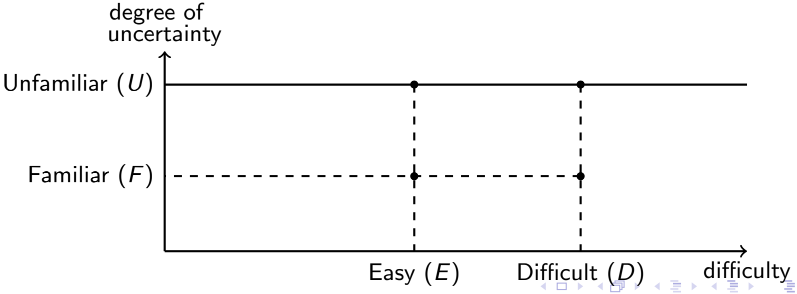
First stage

- A panel with colored balls (red and blue) is drawn
  - Easy panel (100 balls): 51 of dominant color / 49 of other color
  - Difficult panel (400 balls): 201 of dominant color / 199 of other color
- This panel is drawn from a pool (all easy or all difficult)
  - Familiar task: 8 panels of one color / 2 panels of other color
  - Unfamiliar task: 5 panels of one color / 5 panels of other color
- Participants see the drawn panel and estimate the dominant color
- Two treatments: High reward (€10) / Low reward (€0.5)



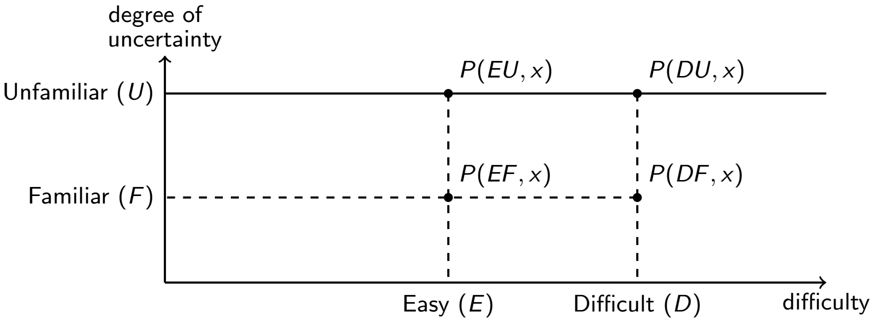
First stage

- A panel with colored balls (red and blue) is drawn
  - Easy panel (100 balls): 51 of dominant color / 49 of other color
  - Difficult panel (400 balls): 201 of dominant color / 199 of other color
- This panel is drawn from a pool (all easy or all difficult)
  - Familiar task: 8 panels of one color / 2 panels of other color
  - Unfamiliar task: 5 panels of one color / 5 panels of other color
- Participants see the drawn panel and estimate the dominant color
- Two treatments: High reward (€10) / Low reward (€0.5)



Main stage: Comparable tasks

Guess: how many participants were correct?



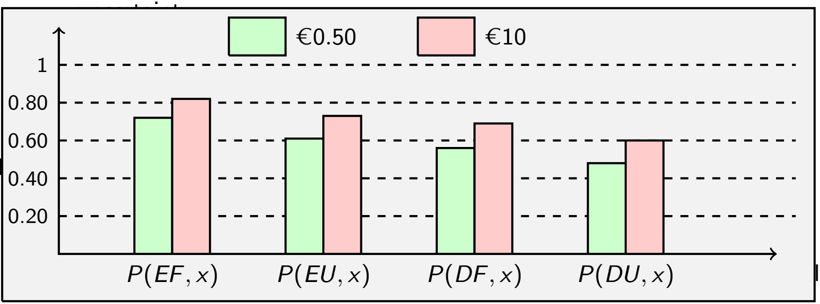
(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:
$P(S, €10) \geq P(S, €0.50)$

(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$
$P(EF,x) \geq P(EU,x) \geq P(DU,x)$ and $P(EF,x) \geq P(DF,x) \geq P(DU,x)$

Main stage: Comparable tasks

Guess: how many participants were correct?

degree of



Un

lty

(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:

$P(S, €10) \geq P(S, €0.50)$

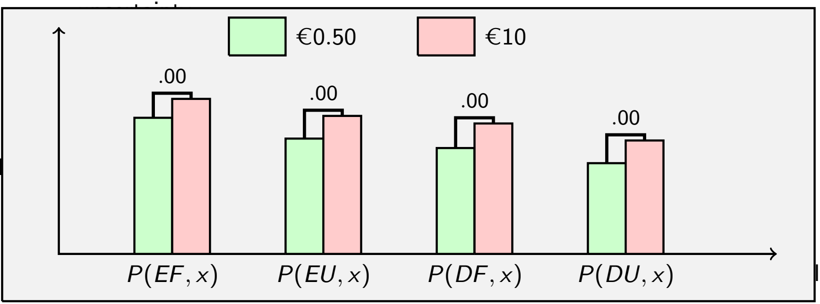
(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$

$P(EF, x) \geq P(EU, x) \geq P(DU, x)$ and $P(EF, x) \geq P(DF, x) \geq P(DU, x)$

Main stage: Comparable tasks

Guess: how many participants were correct?

degree of



(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:

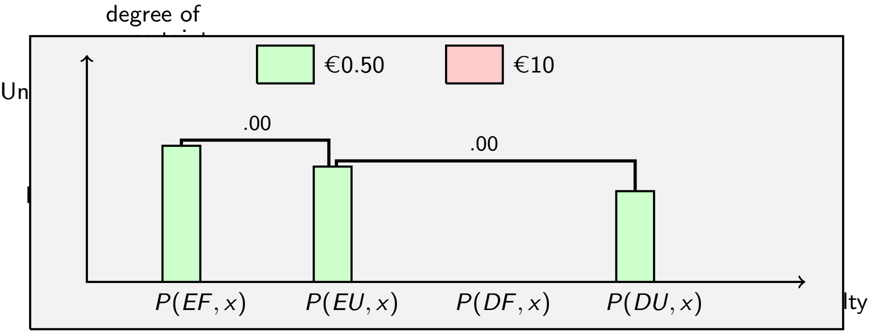
$P(S, €10) \geq P(S, €0.50)$

(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$

$P(EF, x) \geq P(EU, x) \geq P(DU, x)$ and $P(EF, x) \geq P(DF, x) \geq P(DU, x)$

Main stage: Comparable tasks

Guess: how many participants were correct?



(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:

$P(S, €10) \geq P(S, €0.50)$

(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$

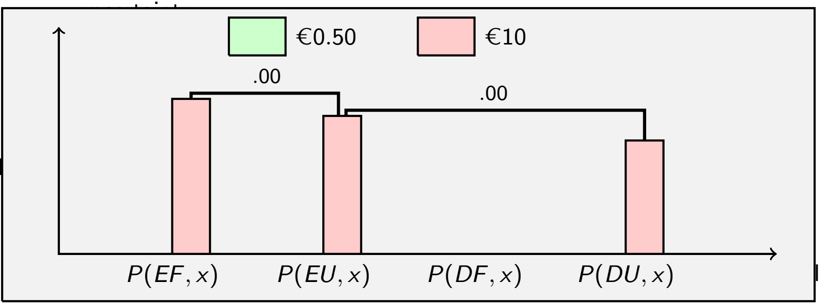
$P(EF, x) \geq P(EU, x) \geq P(DU, x)$ and $P(EF, x) \geq P(DF, x) \geq P(DU, x)$

Main stage: Comparable tasks

Guess: how many participants were correct?

degree of

Un



lty

(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:

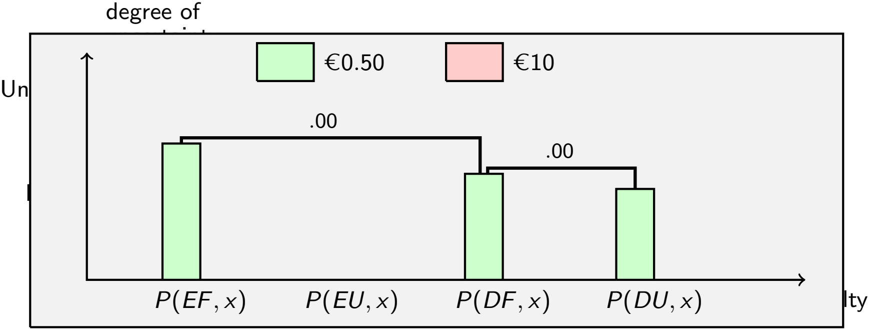
$P(S, €10) \geq P(S, €0.50)$

(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$

$P(EF, x) \geq P(EU, x) \geq P(DU, x)$ and $P(EF, x) \geq P(DF, x) \geq P(DU, x)$

Main stage: Comparable tasks

Guess: how many participants were correct?



(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:

$P(S, €10) \geq P(S, €0.50)$

(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$

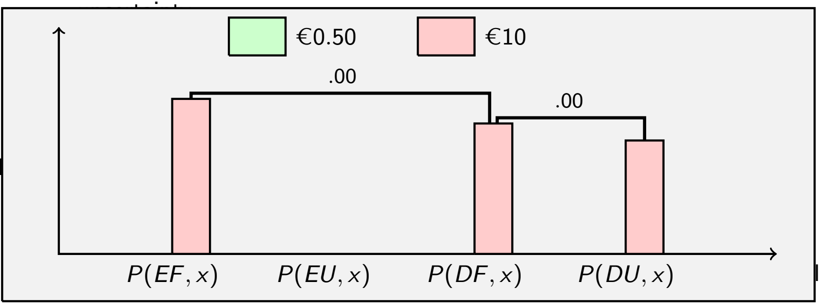
$P(EF, x) \geq P(EU, x) \geq P(DU, x)$ and $P(EF, x) \geq P(DF, x) \geq P(DU, x)$

Main stage: Comparable tasks

Guess: how many participants were correct?

degree of

Un



lty

(H_0) **Sanity check:** For all $S \in \{EU, DU, EF, DF\}$:

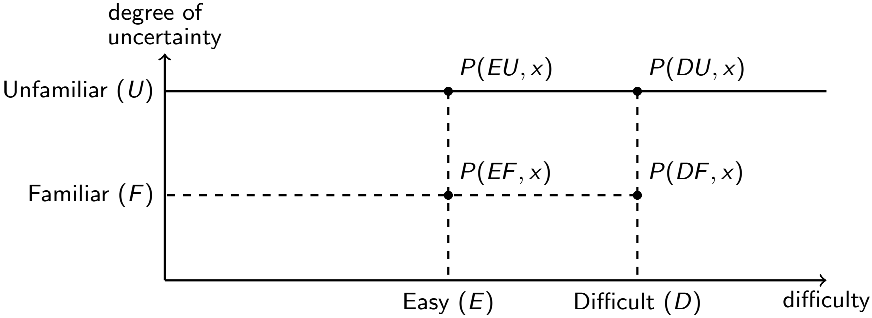
$P(S, €10) \geq P(S, €0.50)$

(H_1) **Basic hypothesis:** For both $x \in \{€0.50, €10\}$

$P(EF, x) \geq P(EU, x) \geq P(DU, x)$ and $P(EF, x) \geq P(DF, x) \geq P(DU, x)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?



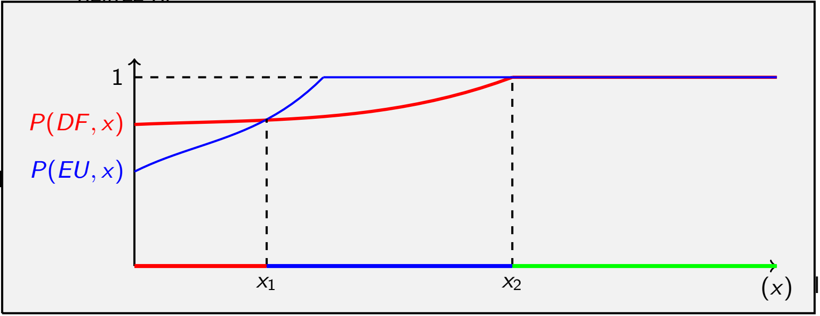
(H_2) **Additional hypotheses:**

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \Rightarrow P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \Rightarrow P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?



(H_2) **Additional hypotheses:**

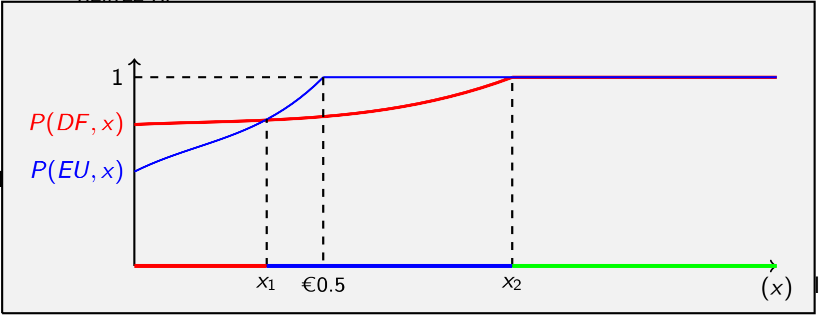
(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \;\Rightarrow\; P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \;\Rightarrow\; P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?

degree of



(H_2) **Additional hypotheses:**

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \Rightarrow P(EU, €10) \geq P(DF, €10)$

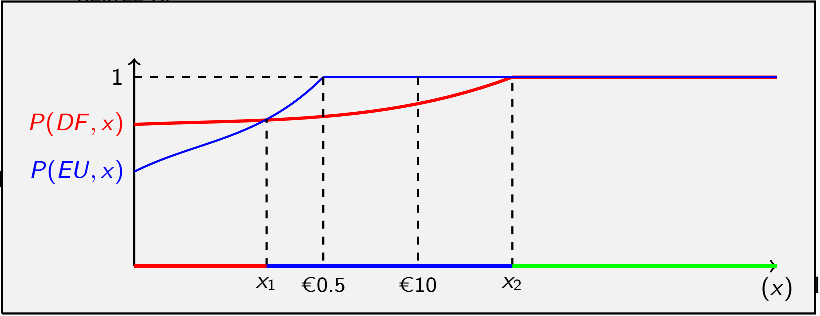
(H_{2b}) $P(EU, €10) \leq P(DF, €10) \Rightarrow P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?

degree of

Un



lty

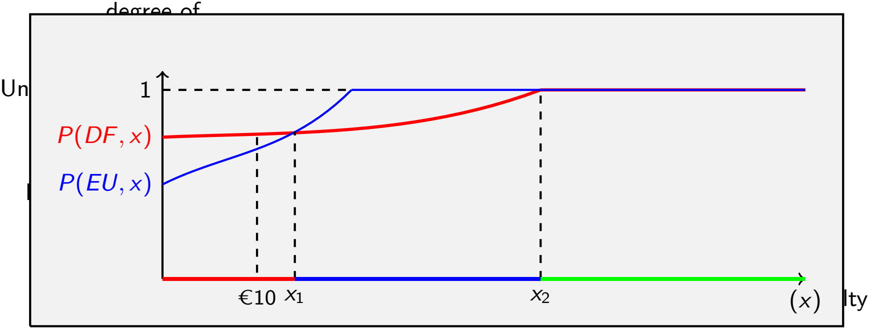
(H_2) **Additional hypotheses:**

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \ \Rightarrow\ P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \ \Rightarrow\ P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?



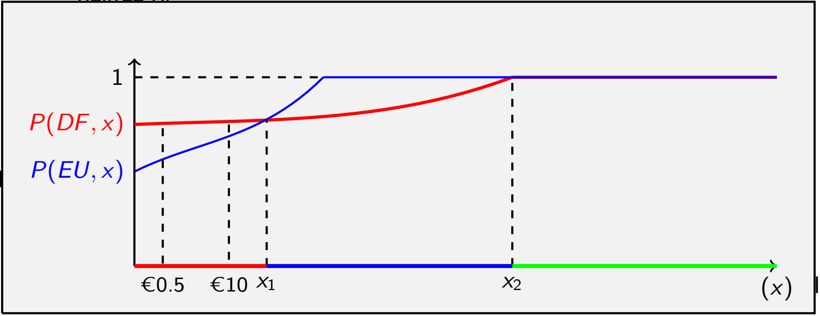
(H_2) **Additional hypotheses:**

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \Rightarrow P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \Rightarrow P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?



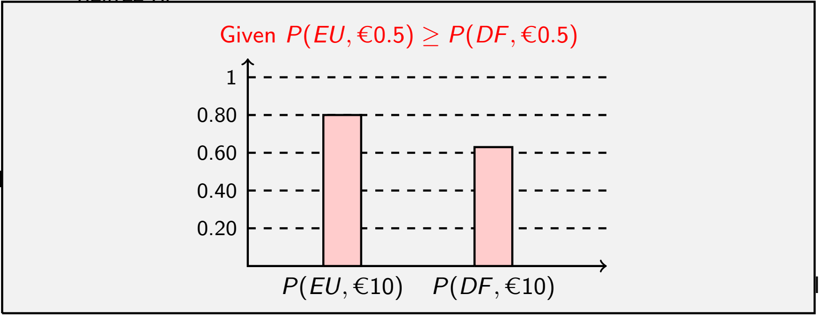
(H_2) Additional hypotheses:

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \Rightarrow P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \Rightarrow P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?



(H_2) **Additional hypotheses:**

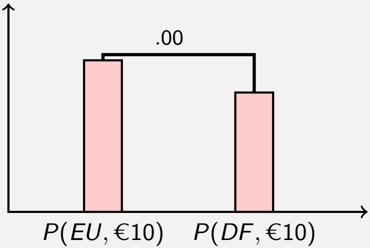
(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \;\Rightarrow\; P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \;\Rightarrow\; P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?

Given $P(EU, €0.5) \geq P(DF, €0.5)$



(H_2) **Additional hypotheses:**

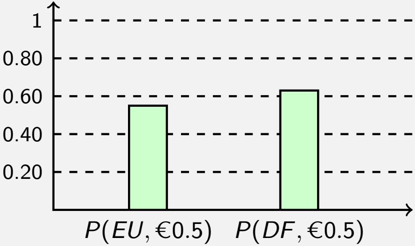
(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \;\Rightarrow\; P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \;\Rightarrow\; P(EU, €0.5) \leq P(DF, €0.5)$

Main stage: Possibly incomparable tasks

Guess: how many participants were correct?

Given $P(EU, €10) \leq P(DF, €10)$



(H_2) **Additional hypotheses:**

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \;\Rightarrow\; P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \;\Rightarrow\; P(EU, €0.5) \leq P(DF, €0.5)$

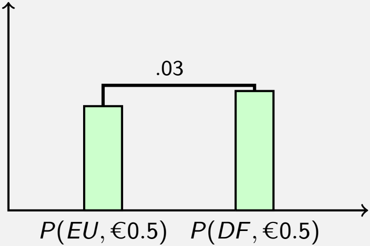
Main stage: Possibly incomparable tasks

Guess: how many participants were correct?

degree of

Un

Given $P(EU, €10) \leq P(DF, €10)$



lty

(H_2) **Additional hypotheses:**

(H_{2a}) $P(EU, €0.5) \geq P(DF, €0.5) \;\Rightarrow\; P(EU, €10) \geq P(DF, €10)$

(H_{2b}) $P(EU, €10) \leq P(DF, €10) \;\Rightarrow\; P(EU, €0.5) \leq P(DF, €0.5)$

Roadmap

Overview of results

1. **Identification**: Task A is more complex than task B if it is both more difficult and less familiar
   - Degree of uncertainty is a novel channel of complexity
   - Difficulty remains the primary channel

2. **Elicitation**: Standard belief elicitation mechanisms reveal whether probability of solving A is larger or smaller than B, even though both might be misreported.
   - Not too difficult to elicit our measure

3. **Validation**: Theoretical predictions corroborated in lab experiment.

4. **Completion (extra result)**: For non-comparable tasks A and B (viz., A is more difficult and more familiar than B), suppose that we start collecting data about B. Then, regardless where the data comes from, eventually A will certainly become more complex than B.
   - We do not need to know anything about the information source

Thanks for your (in)attention ☺