Revealing Complexity

Duarte Gonçalves
University College London

Julen Zarate-Pina
University of the Basque Country

Salvatore Nunnari
Bocconi University

Behavioral/Experimental Seminar

University of Pittsburgh

17 April 2026

Mistakes and Complexity

Why do we care?

Dominated choices: lottery choice, dominant strategy mechanisms,

health insurance plans, pension plans, mortgages, etc.

Cognition and Complexity

A leading explanation for behavioural 'biases':

Cognitive limitations/costs and problem complexity.

(Luce, Wilcox, Rubinstein, Spiegler, Frydman, Jin, Bossaerts, Oprea, Enke, Zimmermann, Graeber, Salant, Agranov, Esponda, Vespa, Yuksel, Martínez-Marquina, Niederle, Puri, Alaoui Penta, ...; also Li, Borgers Li, Pycia Troyan, Camara, ...)

+ Complex problems $\rightarrow$ + Strain on cognitive resources $\rightarrow$ + Mistakes.

More complex problem if more mistakes and/despite greater cognitive effort.

Mistakes and Complexity

Why do we care?

Dominated choices: lottery choice, dominant strategy mechanisms,

health insurance plans, pension plans, mortgages, etc.

Cognition and Complexity

A leading explanation for behavioural 'biases':

Cognitive limitations/costs and problem complexity.

(Luce, Wilcox, Rubinstein, Spiegler, Frydman, Jin, Bossaerts, Oprea, Enke, Zimmermann, Graeber, Salant, Agranov, Esponda, Vespa, Yuksel, Martínez-Marquina, Niederle, Puri, Alaoui Penta, ...; also Li, Borgers Li, Pycia Troyan, Camara, ...)

+ Complex problems $\to$ + Strain on cognitive resources $\to$ + Mistakes.

More complex problem if more mistakes and/despite greater cognitive effort.

− Ability $\to$ + Cognitive costs $\to$ + Mistakes.

Less able agents treat same problem as if more complex.

Mistakes, Complexity and Ability

Inferring complexity and ability

Understand when decision problem challenging to make it simpler (or not).

Select agents with higher ability.

Mistakes, Complexity and Ability

Inferring complexity and ability

Understand when decision problem challenging to make it simpler (or not).

Select agents with higher ability.

Problem: Often no measures accuracy available.

Want method to infer complexity based on observable behaviour.

Mistakes, Complexity and Ability

Inferring complexity and ability

- Understand when decision problem challenging to make it simpler (or not).
- Select agents with higher ability.

Problem: Often no measures accuracy available.

- Want method to infer complexity based on observable behaviour.

This paper: Experimentally test a method to infer problem complexity and agent ability from choices alone.

- Applied to wide range of problem domains: perception, computation, deduction, prediction, inference.
- Based on sequential sampling framework.
- Predictions involving choices and response time supported by the data.
- Response time not good proxy for complexity.
- Importance of meta-cognition and learning (a problem's) complexity.

Overview

1. Why Not Just Use Response Times?
2. A Method to Infer Complexity (And Ability)
3. Experimental Design
4. Prediction 1: Accuracy and Time
5. Prediction 2: Effect of Subsidies on Choices and Time
6. Prediction 3: Revealing Complexity
7. Prediction 4: Revealing Ability
8. Meta-Cognition and Learning
9. Prediction 0: Exogenous Stopping
10. Related Literature
11. Conclusion

Why Not Just Use Response Times?

Mistakes and Complexity

Inferring complexity

Understand when decision problem challenging and make it simpler (or not).

Mistakes and Complexity

Inferring complexity

Understand when decision problem challenging and make it simpler (or not).

Usual proxy: **response times**

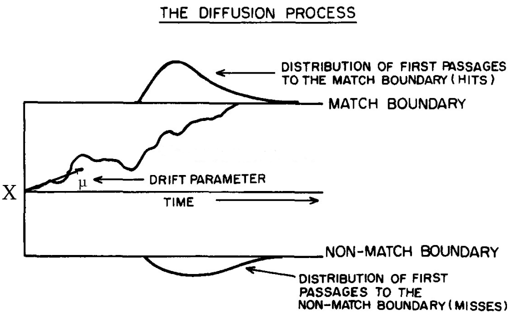
(Stroop 35; Hick 52; Shepard Metzler 71; Treisman Gelade 80; Bassili Scott 96; Roitman Shadlen 02; Wilson et al. 10; Murawski Bossaerts 16; Franco et al. 21; Gill Prowse 23; Hong Stauffer 23; ...)

Easy choices will produce fast and accurate responses, while difficult ones will be time consuming and poorly efficient

(Cerreia-Vioglio Maccheroni Marinacci Rustichini 22)

Faster and better is easier; Slower and worse is more complex.

Speed-Accuracy and the Drift-Diffusion Model



(adapted from Ratcliff 78)

Drift-Diffusion Model. Ratcliff (1978): influential paper in cognitive sciences, 6k+ cites.

Model noisy cognition relating time and choice via sequential sampling.

Exogenous stopping: stop whenever sufficiently convinced.

Prediction 0

Under exogenous stopping, accuracy is monotone in problem complexity, whereas time is single-peaked.

(All predictions from Gonçalves, *Speed, Accuracy, and Complexy*.)

From Faster is Better to Faster is Easier

Speed-Accuracy Trade-off

More time, more information: better choices, but more costly.

Less decisive info $\rightarrow$ Slower and worse choices. **Slower is worse.**

Slower and worse is more complex.

$\Longrightarrow$ **Slower is more complex.**

Complex Problem $\rightarrow$ Noisier Cognition $\rightarrow$ Weaker Signal $\rightarrow$ Lower Accuracy & Slower.

Similar reasoning for ability.

Speed, Accuracy and Complexity

Runs against intuition: **Faster can't always be easier and better**.

- If too complicated, not going to try too hard.

Speed, Accuracy and Complexity

Runs against intuition: **Faster can't always be easier and better**.

If too complicated, not going to try too hard.

From Exogenous to Optimal Stopping

Optimal stopping: expect DM to trade-off accuracy and effort/time.

Behaviour depends on incentives and problem complexity.

"The assumption is that payoffs or instructions induce the subject to adjust the positions of the [...] boundaries and so to adjust the amount of information required for a decision" (Ratcliff 78)

Speed, Accuracy and Complexity

Prediction 1

Under optimal stopping, accuracy is monotone in problem complexity, whereas time is single-peaked.

Importance of being aware: adjusting vs not adjusting to complexity.

- Not adjusted/exogenous: diff. in expected payment across health insurance plans.
- Adjusted: comparison over plans with more/fewer contractual clauses.

Speed, Accuracy, and Ability

Response time also standard proxy for ability.

Slower = more able: oftentimes documented rather than assumed relationship.

Intuition: + ability $\rightarrow$ + effort $\rightarrow$ slower.

financial choices (Darriet et al. 20); dominance-solvable games (Kocher Sutter 06; Rubinstein 07, 16; Agranov Caplin Tergiman 15; Alós-Ferrer Buckenmaier 21; Esteban-Casanelles Gonçalves WP; Gill Prowse 23); public goods (Recalde Riedl Vesterlund 18).

Speed, Accuracy, and Ability

Response time also standard proxy for ability.

Slower = more able: oftentimes documented rather than assumed relationship.

Intuition: + ability $\rightarrow$ + effort $\rightarrow$ slower.

financial choices (Darriet et al. 20); dominance-solvable games (Kocher Sutter 06; Rubinstein 07, 16; Agranov Caplin Tergiman 15; Alós-Ferrer Buckenmaier 21; Esteban-Casanelles Gonçalves WP; Gill Prowse 23); public goods (Recalde Riedl Vesterlund 18).

Faster = more able: huge literature in psychology dating back to Thorndike, Bregman, Cobb, and Woodyard 1926 (e.g., Jensen Munro 79; Jensen 81, 93; Goldhammer 15).

other things being equal, if intellect A can do at each level the same number of tasks as intellect B, but in a less time, intellect A is better.

Speed, Accuracy, and Ability

Prediction 1'

Under optimal stopping, accuracy is monotone in ability, whereas time is single-peaked. Furthermore, time has single-crossing property in ability and problem complexity.

Single-crossing:

in simple problems, higher ability chooses faster and better;

in complex problems, higher ability chooses slower and better.

Recent evidence from psychology and education research on non-monotonicity of time.

(Lindley et al. 95; Dodonova Dodonov 13; Goldhammer et al. 15)

Also in lottery choice (Agranov Trevino Schotter WP).

A Method to Infer Complexity (And Ability)

Inferring Complexity

Back to inferring complexity...

Inferring Complexity

Back to inferring complexity...

There is a correct answer (and analyst knows it).

Choices tend to identify complexity:

$\rightarrow$ (avg) accuracy decreasing in problem complexity.

Inferring Complexity

Back to inferring complexity...

There is a correct answer (and analyst knows it).

Choices tend to identify complexity:

$\rightarrow$ (avg) accuracy decreasing in problem complexity.

There is no correct answer/Analyst doesn't know it.

E.g., buying products, reporting preferences, subjective beliefs.

Response time often used as proxy, but not always appropriate:

Non-monotone relationship creates inference problem.

Going to higher moments doesn't help.

What to do?

Inferring Complexity

Back to inferring complexity...

There is a correct answer (and analyst knows it).

Choices tend to identify complexity:

$\rightarrow$ (avg) accuracy decreasing in problem complexity.

There is no correct answer/Analyst doesn't know it.

E.g., buying products, reporting preferences, subjective beliefs.

Response time often used as proxy, but not always appropriate:

Non-monotone relationship creates inference problem.

Going to higher moments doesn't help.

What to do?

Subsidise alternative.

Incentives and Time

Prediction 2

Under optimal stopping, subsidising an alternative leads to it being chosen more often and faster, and to choosing other alternatives less often and slower.

Intuition: Higher subsidy $\rightarrow$ need to be less convinced it's good to choose it.

Robust feature of sequential sampling.

Incentives and Complexity

Prediction 3

Under optimal stopping, choices are more responsive to subsidies in more complex problems.

Intuition: If problem simple enough, then close to sure,

1 penny more won't affect choices much.

If very complex, subsidy overwhelmingly affects choices.

Incentives and Complexity

Prediction 3

Under optimal stopping, choices are more responsive to subsidies in more complex problems.

Intuition: If problem simple enough, then close to sure,

1 penny more won't affect choices much.

If very complex, subsidy overwhelmingly affects choices.

Important: need to recognise complexity; meta-cognition.

Incentives and Complexity

Prediction 3

Under optimal stopping, choices are more responsive to subsidies in more complex problems.

Intuition: If problem simple enough, then close to sure,
1 penny more won't affect choices much.

If very complex, subsidy overwhelmingly affects choices.

Important: need to recognise complexity; meta-cognition.

Prediction 4

Under optimal stopping, choices are more responsive to subsidies for DMs with lower ability.

Intuition: Same problem seems more complex to less able DM.

Experimental Design

Experimental Design

Task Domains

Task Domains

Test predictions in a systematic manner across different domains, appealing to different cognitive processes.

Perception

- Large literature on perceptual decision-making and sensory-signal processing in psychology and neuroscience (Summerfield Blangero 17; Najafi Churchland 18).
- Other applications: advertising and marketing, representation of financial information, regulation and product labelling, tax salience, etc.
- ID people over/under 21 (Caplin et al. WP), food labelling (Ravaioli WP; Barahona et al. 23)

Task Domains

Test predictions in a systematic manner across different domains, appealing to different cognitive processes.

Perception, **Computation**

Computational complexity underlies suboptimal reaction to tax schedules (Rees-Jones Taubinsky 20), implementing rules (Oprea 20), identifying causal relations (Kendall Oprea 24), knapsack problems (Bossaerts Murawski 17), consumer search and attribute aggregation (Sanjurjo WP), exponential growth bias and suboptimal financial choices (Wagenaar Sagaria 75; Eisenstein Hoch 07; Levy Tasoff 16; Carvalho Silverman 24).

Task Domains

Test predictions in a systematic manner across different domains, appealing to different cognitive processes.

Perception, Computation, **Logic**

- Logical reasoning is longstanding topic studied in psychology (Johnson-Laird 72, 83, 92, 06; Copeland Radvansky 04).
- Propositional knowledge underpins standard knowledge model (Aumann 98).
- Logical complexity relevant for contract specification and take-up (Jakobsen 20), bank regulation design (Colliard Georg WP), and failures of contingent reasoning — e.g., voting, college admissions, multi-object allocation, health insurance (Niederle Vespa 23; Nagel Serrano WP).

Task Domains

Test predictions in a systematic manner across different domains, appealing to different cognitive processes.

Perception, Computation, Logic, **Prediction**

- Biased predictions are major topic of interest in macroeconomics (Coibion Gorodnichenko 12; Bordalo Gennaioli Ma Shleifer 20; Afrouzi Kwon Landier Ma Thesmar 23; Fan Liang Peng WP).
- Complex DGPs raise difficulty of understanding underlying causal structure and adopting correct mental model (Esponda Vespa Yuksel 23; Oprea 20; Kendall Oprea 24).

Task Domains

Test predictions in a systematic manner across different domains, appealing to different cognitive processes.

Perception, Computation, Logic, Prediction, **Inference**.

- Extensive literature on belief updating in economics, psychology, and political science (Benjamin 19).
- Belief updating from info crucial for most day-to-day situations: understanding recommendations and reviews [1]Aridor Gonçalves et al. 24, WP), reacting to public and private communitation, correcting misinformation [2]Nyhan et al. 19).
- Complexity of updating (Gonçalves et al. 25; Agranov Reshidi WP).

Deploying the Method

Focus on binary choice: A vs B.

Subsidy to A vs Subsidy to B.

E/not E: domain-specific condition.

- Option A: £subsidy a for sure + £x if E
- Option B: £subsidy b for sure + £x if not E

Vary subsidies to reveal complexity/ability.

Experimental Design

Perception

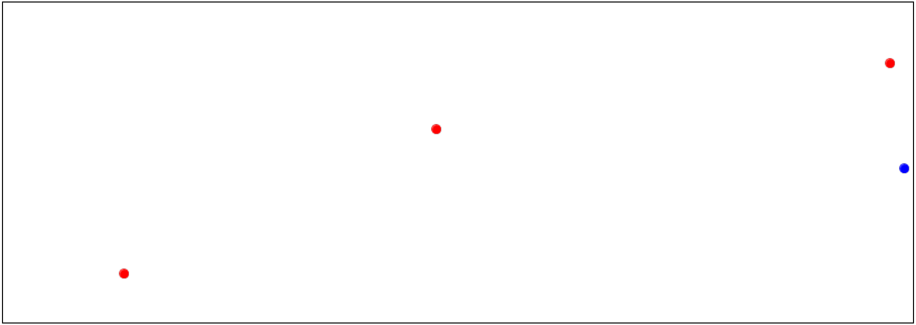
Perception

Round 3

Which option do you choose?

There are 4 dots: 3 of one colour and 1 of the other colour.

- ◯ Option A: **£0.75** for sure + **£2.00** if most dots are Blue
- ◯ Option B: **£0.25** for sure + **£2.00** if most dots are Red



Next

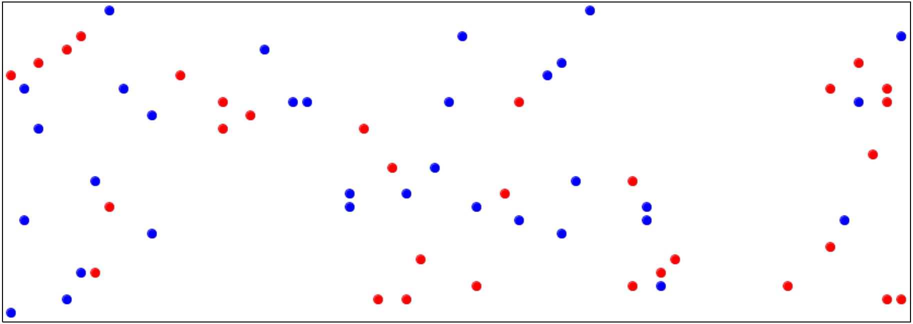
Show Instructions

Round 2

Which option do you choose?

There are 64 dots: 33 of one colour and 31 of the other colour.

- Option A: **£0.75** for sure + **£2.00** if most dots are Blue
- Option B: **£0.25** for sure + **£2.00** if most dots are Red



Next

Show Instructions

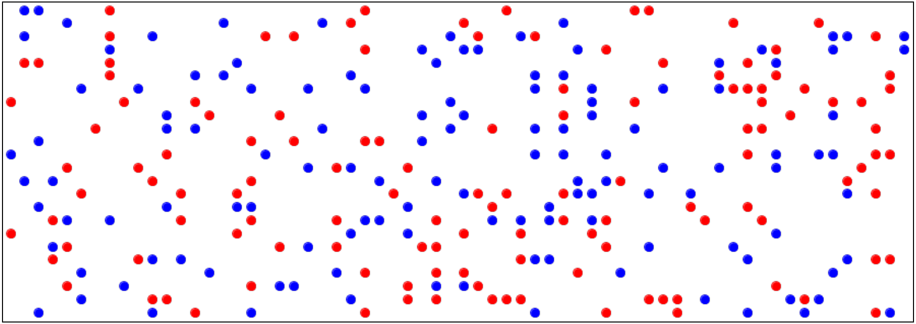
Perception

Round 6

Which option do you choose?

There are 256 dots: 129 of one colour and 127 of the other colour.

- Option A: **£0.25** for sure + **£2.00** if most dots are Blue
- Option B: **£0.25** for sure + **£2.00** if most dots are Red



Next

Show Instructions

Perception

Standard task paradigm for study of perceptual discrimination, dating back to at 1940s

(Philips 47; in economics, Caplin & Dean 13).

E: most dots are Blue/Red.

Variations:

Number of balls: 4, 16, 64, 256.

Randomise majority colour and position.

Experimental Design

Computation

Round 2

Which option do you choose?

Each sum has 2 numbers.

- ○ Option A: **£0.25** for sure + **£2.00** if Sum 1 has the largest value
- ○ Option B: **£0.25** for sure + **£2.00** if Sum 2 has the largest value

Sum 1	**Sum 2**
four minus five	nine plus two

Next

Show Instructions

Round 4

Which option do you choose?

Each sum has 8 numbers.

- Option A: **£0.75** for sure + **£2.00** if Sum 1 has the largest value
- Option B: **£0.25** for sure + **£2.00** if Sum 2 has the largest value

Sum 1	**Sum 2**
seven minus four minus zero plus two plus zero minus seven minus five minus four	two plus six minus eight minus six minus two minus five minus seven plus zero

Next

Show Instructions

Round 3

Which option do you choose?

Each sum has 64 numbers.

- Option A: **£0.25** for sure + **£2.00** if Sum 1 has the largest value
- Option B: **£0.75** for sure + **£2.00** if Sum 2 has the largest value

Sum 1

nine minus one minus seven minus eight minus three plus zero plus two plus seven minus five minus nine minus two minus five plus zero plus five minus four plus four minus six plus two minus four minus four minus five minus five minus nine plus one minus seven plus two minus eight plus one minus three minus one minus two plus three minus one minus four minus zero plus seven plus six minus two minus three plus eight plus eight minus six plus four minus seven minus eight minus six minus three plus three plus seven plus zero minus three minus five plus eight minus two plus three minus zero plus eight minus

Sum 2

four minus seven minus nine minus eight minus three plus three plus seven minus one minus one minus five plus six minus eight minus four minus one plus nine minus nine plus four plus seven minus zero minus nine plus zero plus three plus eight minus zero plus eight plus four minus nine plus four minus six minus one plus four plus seven minus zero plus nine minus eight minus two minus three minus four plus zero minus eight plus nine plus eight plus two minus six plus two plus zero minus three plus seven minus eight plus two minus eight minus seven plus one plus seven minus three minus eight minus eight

Computation

Standard task paradigm for study of arithmetic cognition

(cf. Ashcraft 92; in economics, Caplin & Martin 11, 15).

E: Sum 1/2 has largest value.

Variations:

Number of summands: 2, 8, 32, 64.

Randomise summands uniformly 0-9, operators plus/minus; no ties.

Experimental Design

Logic

Logic

Logical Statements:

New task inspired by work by Johnson-Laird, Copeland & Radvansky...
and Agatha Christie.

List of suspects, each equally likely to be guilty or innocent.

List of statements, clues to solving the case.

E: suspect X is guilty/innocent.

Variations:

Number of statements: 2, 3, 16.

Other variations: randomly selected suspect X, type and position of statements (atom, if A then B, A iff not B, etc.), underlying Agatha Christie characters.

Round 1

Which option do you choose?

- ◯ Option A: **£0.25** for sure + **£2.00** if Guido is innocent
- ◯ Option B: **£0.25** for sure + **£2.00** if Guido is guilty

The suspects are: Andrew and Guido.
Each suspect is equally likely to be guilty or innocent; there may be multiple guilty and innocent suspects.

These are the clues to solve the case:

1. Andrew is guilty.
2. If Guido is guilty, then Andrew is innocent.

Round 2

Which option do you choose?

- ◯ Option A: **£0.75** for sure + **£2.00** if Wilfred is innocent
- ◯ Option B: **£0.25** for sure + **£2.00** if Wilfred is guilty

The suspects are: Amanda, Marlene, and Wilfred.
Each suspect is equally likely to be guilty or innocent; there may be multiple guilty and innocent suspects.

These are the clues to solve the case:

1. Either Marlene and Wilfred are both innocent, or both guilty.
2. Marlene is guilty.
3. If Amanda is innocent, then Wilfred is innocent.

Round 3

Which option do you choose?

- ○ Option A: **£0.75** for sure + **£2.00** if Samuel is innocent
- ○ Option B: **£0.25** for sure + **£2.00** if Samuel is guilty

The suspects are: Ephraim, Roderick, and Samuel.
Each suspect is equally likely to be guilty or innocent; there may be multiple guilty and innocent suspects.

These are the clues to solve the case:

1. If Roderick is guilty, then Samuel is innocent.
2. Either Roderick and Ephraim are both innocent or both guilty.
3. If Ephraim is innocent, then Samuel and Roderick are guilty.

Round 9

Which option do you choose?

- ◯ Option A: **£0.25** for sure + **£2.00** if Leonie is innocent
- ◯ Option B: **£0.75** for sure + **£2.00** if Leonie is guilty

The suspects are: Giles, Helen, Jackie, James, Janet, Joan, Leonie, Lily, Manning, Mrs Cocker, Richard, and the Major.
Each suspect is equally likely to be guilty or innocent; there may be multiple guilty and innocent suspects.

These are the clues to solve the case:

1. Either Manning is guilty or Janet is innocent (or both).
2. If Richard is guilty, then Lily and James are both innocent.
3. If Helen is guilty, then Jackie is innocent.
4. Giles and James are either both guilty or both innocent.
5. Either James is guilty or Leonie is, but not both.
6. If Giles is innocent, then Helen and the Major are both guilty.
7. If Joan is guilty, then the Major is also guilty.
8. Either Lily and Joan are both guilty or both innocent.
9. If Richard is innocent, then so is Leonie.
10. If Jackie is guilty, then Manning is also guilty.
11. If Manning is innocent, then Helen is guilty.
12. If the Major is guilty, then Manning and Mrs Cocker are innocent.
13. If Mrs Cocker is guilty, then Richard and Jackie are both innocent.
14. Janet is guilty if and only if the Major is innocent.
15. Janet and Giles cannot both be guilty.
16. Either Lily is guilty and Janet is innocent, or vice versa.

Experimental Design

Prediction

Prediction

Automata:

DGP depends on initial state and exogenous inputs, x/y, to produce outputs, R/G.

Observe: sequence of inputs and outputs.

Goal: predict next output.

Task based on Kendall & Oprea (24). Main difference: we describe automaton.

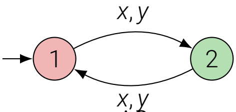
E: next output is R/G.

Variations:

Automata: A2, I2, S2, H8.

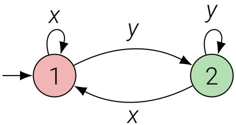
Other variations: random initial state, inputs.

Two-State Autonomous
A fixed pattern,
independent of inputs



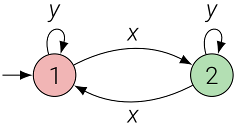
A2

Two-State Instruct
Input directly
determines state



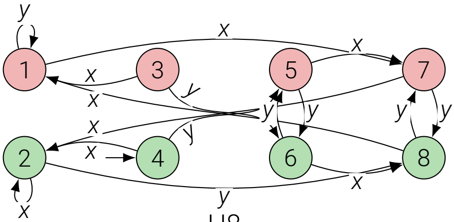
I2

Two-State Switch
One input keeps,
the other toggles state



S2

Eight-state Hybrid
More complex pattern
over eight states



H8

Prediction: A2

Round 5

Which option do you choose?

- ◯ Option A: **£0.25** for sure + **£2.00** if R is next in the sequence
- ◯ Option B: **£0.25** for sure + **£2.00** if G is next in the sequence

This is how outputs are generated:

The system can be in one of two states, 1 and 2.
If the previous state is 1 and the input is x, then the state is 2.
If the previous state is 1 and the input is y, then the state is 2.
If the previous state is 2 and the input is x, then the state is 1.
If the previous state is 2 and the input is y, then the state is 1.

This is how the state is determined:

The initial state is 1.
If the state is 1, then the output is R.
If the state is 2, then the output is G.

You only see the current input and all past outputs and inputs, but not the state.

												↓
Inputs		y	x	y	x	x	y	x	y	x	x	y
Outputs	R	G	R	G	R	G	R	G	R	G	R	**?**

Round 6

Which option do you choose?

- ◯ Option A: **£0.25** for sure + **£2.00** if G is next in the sequence
- ◯ Option B: **£0.25** for sure + **£2.00** if R is next in the sequence

This is how outputs are generated:

The system can be in one of two states, 1 and 2.
If the previous state is 1 and the input is y, then the state is 2.
If the previous state is 1 and the input is x, then the state is 1.
If the previous state is 2 and the input is y, then the state is 2.
If the previous state is 2 and the input is x, then the state is 1.

This is how the state is determined:

The initial state is 2.
If the state is 1, then the output is G.
If the state is 2, then the output is R.

You only see the current input and all past outputs and inputs, but not the state.

Inputs		x	y	x	x	y	x	y	x	y	x	x
Outputs	R	G	R	G	G	R	G	R	G	R	G	**?**

Prediction: S2

Round 4

Which option do you choose?

- Option A: **£0.25** for sure + **£2.00** if R is next in the sequence
- Option B: **£0.75** for sure + **£2.00** if G is next in the sequence

This is how outputs are generated:

The system can be in one of two states, 1 and 2.
If the previous state is 1 and the input is x, then the state is 2.
If the previous state is 1 and the input is y, then the state is 1.
If the previous state is 2 and the input is x, then the state is 1.
If the previous state is 2 and the input is y, then the state is 2.

This is how the state is determined:

The initial state is 2.
If the state is 1, then the output is R.
If the state is 2, then the output is G.

You only see the current input and all past outputs and inputs, but not the state.

Inputs		y	y	x	y	y	y	y	x	x	x	y
Outputs	G	G	G	R	R	R	R	R	G	R	G	**?**

Prediction: H8

Round 1

Which option do you choose?

- Option A: **£0.25** for sure + **£2.00** if R is next in the sequence
- Option B: **£0.75** for sure + **£2.00** if G is next in the sequence

This is how outputs are generated:

The system can be in one of eight states: 1, 2, 3, 4, 5, 6, 7, 8.
If the previous state is 1 and the input is x, then the state is 7.
If the previous state is 1 and the input is y, then the state is 1.
If the previous state is 2 and the input is x, then the state is 2.
If the previous state is 2 and the input is y, then the state is 8.
If the previous state is 3 and the input is x, then the state is 1.
If the previous state is 3 and the input is y, then the state is 6.
If the previous state is 4 and the input is x, then the state is 2.
If the previous state is 4 and the input is y, then the state is 5.
If the previous state is 5 and the input is x, then the state is 7.
If the previous state is 5 and the input is y, then the state is 6.
If the previous state is 6 and the input is x, then the state is 8.
If the previous state is 6 and the input is y, then the state is 5.
If the previous state is 7 and the input is x, then the state is 2.
If the previous state is 7 and the input is y, then the state is 8.
If the previous state is 8 and the input is x, then the state is 1.
If the previous state is 8 and the input is y, then the state is 7.

This is how the state is determined:

The initial state is 4.
If the state is 1 or 3 or 5 or 7, then the output is R.
If the state is 2 or 4 or 6 or 8, then the output is G.

You only see the current input and all past outputs and inputs, but not the state.

Experimental Design

Other Experimental Details

Experimental Design

Inference

Inference

Balls-and-urns paradigm (Edwards & Phillips 64; cf. Benjamin 19).

Two urns with different composition. Balls drawn. (Details)

E: urn 1/2 is more likely.

Variations:

(i) Informativeness (Blackwell) experiment: different pairs = % reduction in prior entropy.

(ii) Signal strength: different log-likelihood of realised draw.

(iii) Symmetric/Asymmetric info structure: AA' and BB' vs AB and AB'.

(iv) Prior Info (2nd draw) vs No Prior Info (1st draw).

(v) Number Signals: 2 vs 4 colours for balls.

(vi) Simultaneous/Sequential signals: 2 draws at once vs 1 signal at the time.

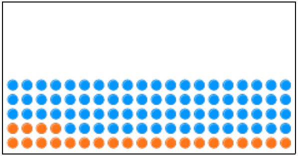
Round 3

Which option do you choose?

- Option A: **£0.25** for sure + **£2.00** if it is more probable that the ball was drawn from Box 1
- Option B: **£1.25** for sure + **£2.00** if it is more probable that the ball was drawn from Box 2

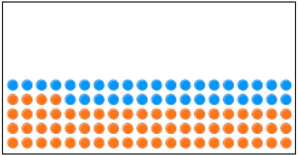
Box 1

Box 1 contains 100 balls, including
24 orange balls and 76 blue balls.



Box 2

Box 2 contains 100 balls, including
64 orange balls and 36 blue balls.



THE BALL DRAW WAS: 

Next

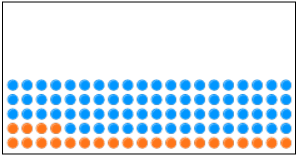
Round 2

Which option do you choose?

- Option A: **£0.25** for sure + **£2.00** if it is more probable that the balls were drawn from Box 1
- Option B: **£1.25** for sure + **£2.00** if it is more probable that the balls were drawn from Box 2

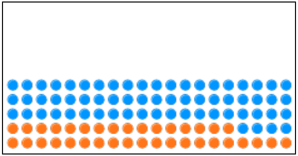
Box 1

Box 1 contains 100 balls, including
24 orange balls and 76 blue balls.



Box 2

Box 2 contains 100 balls, including
36 orange balls and 64 blue balls.



THE BALL DRAWS WERE: 

(Recall after being drawn each ball is put back in the box, and so the same ball can be drawn multiple times.)

Next

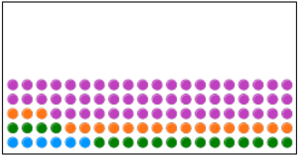
Round 1

Which option do you choose?

- Option A: **£0.25** for sure + **£2.00** if it is more probable that the balls were drawn from Box 1
- Option B: **£1.25** for sure + **£2.00** if it is more probable that the balls were drawn from Box 2

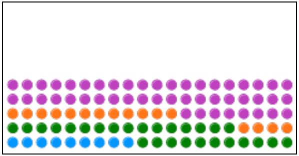
Box 1

Box 1 contains 100 balls, including
6 blue balls, 18 green balls,
19 orange balls, and 57 purple balls.



Box 2

Box 2 contains 100 balls, including
9 blue balls, 27 green balls,
16 orange balls, and 48 purple balls.



THE BALL DRAWS WERE: 

(Recall after being drawn each ball is put back in the box, and so the same ball can be drawn multiple times.)

Next

Other Experimental Details

Participants: 1,590; Prolific, UK-based.

Perception: 235; Computation: 214; Prediction: 300; Logic: 300; Inference: 541.

Rounds: Perception: 48; Computation: 48; Prediction: 36; Logic: 36; Inference: 50.
+2 Practice Rounds.

Payments: pay one round randomly selected.
Avg payment $\approx$ £15-17.50/hour. Avg duration $\approx$ 20-36min.

Bonus conditional on E: £2.00.

Subsidies: allocated independently from E.

Perception and Computation: 0.25 vs 0.25, 0.75, 1.25.

Logic, and Prediction: 0.25 vs 0.25, 0.75.

Inference: 0.25 vs 0.25, 1.25.

Comprehension questions; cannot copy.

Pre-registered on AsPredicted.

Prediction 1: Accuracy and Time

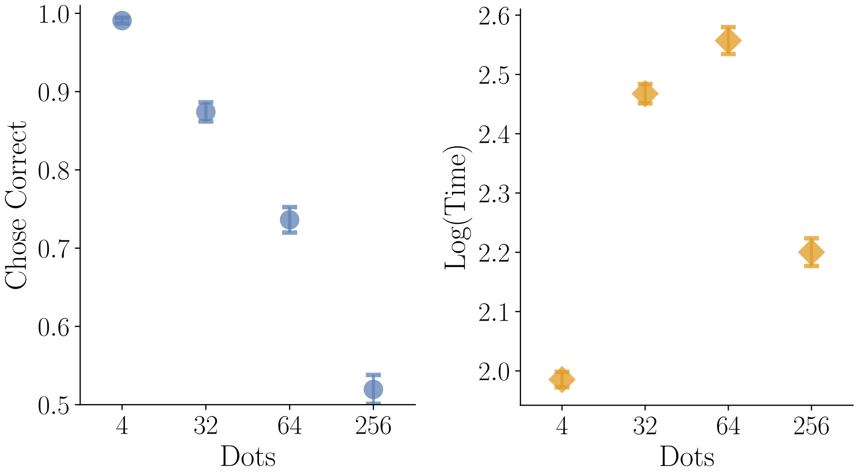
Prediction 1: Accuracy and Time

Accuracy is monotone in problem complexity; time is single-peaked.

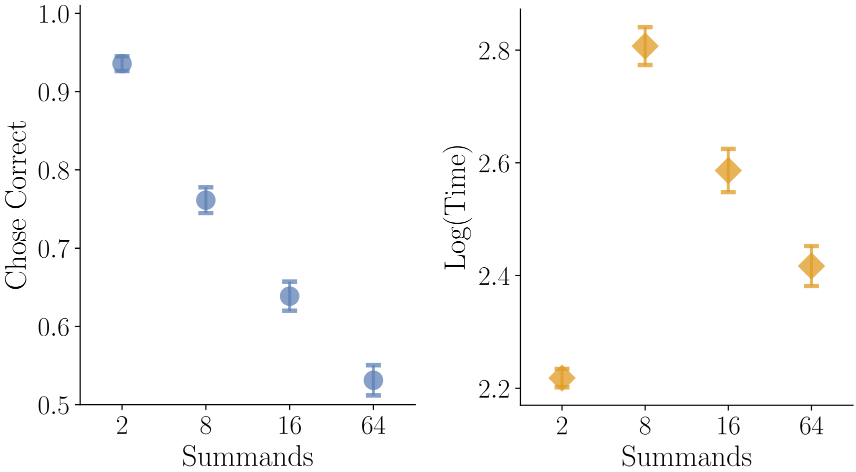
Testing Prediction:

Regression analysis: here boils down to comparing means by condition.

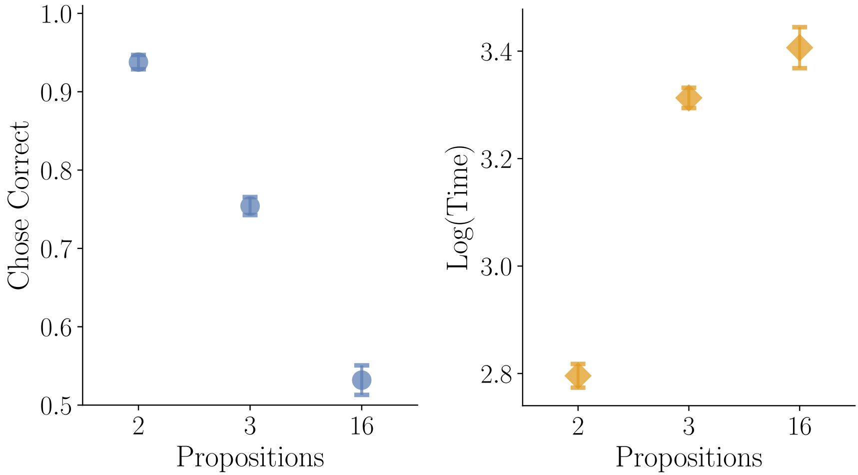
Prediction 1: Accuracy and Time. Perception



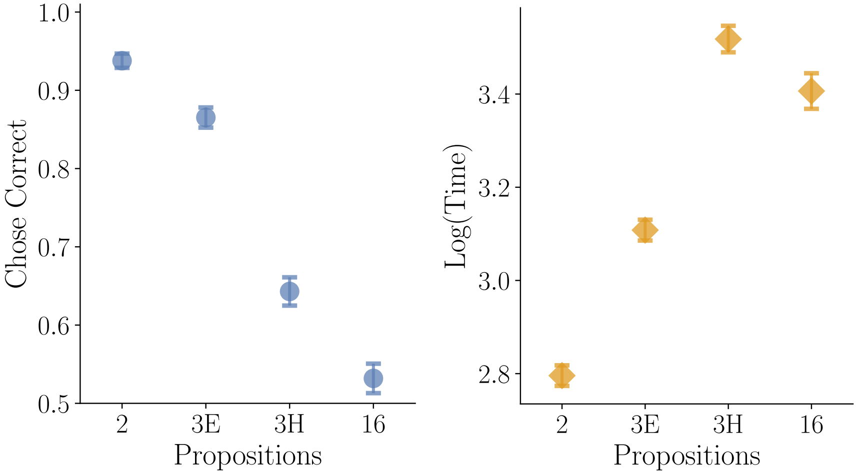
Prediction 1: Accuracy and Time. Computation



Prediction 1: Accuracy and Time. Logic

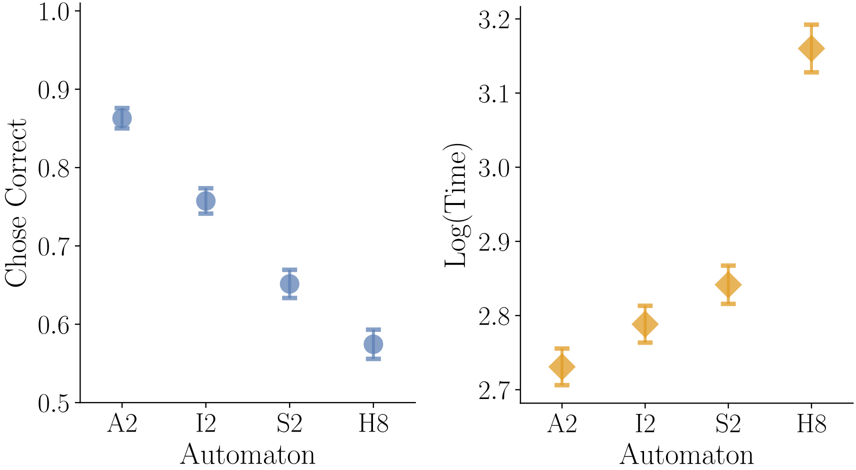


Prediction 1: Accuracy and Time. Logic

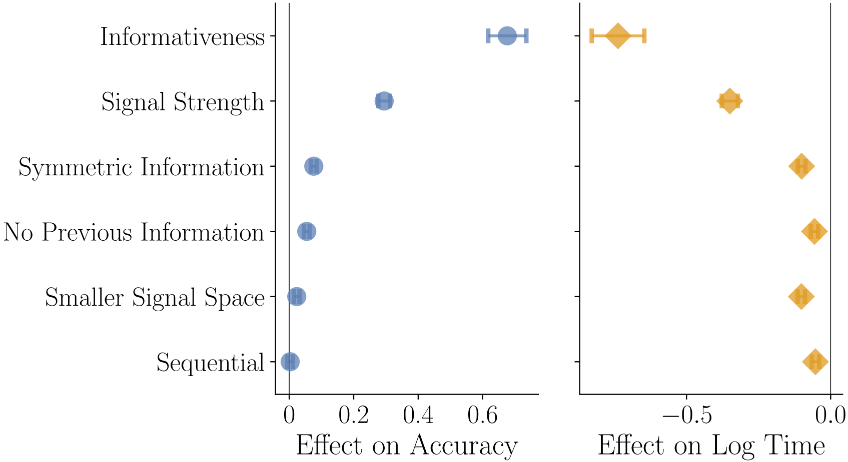


3 Easy/Hard: presence of atom (X is Guilty/Innocent).

Prediction 1: Accuracy and Time. Prediction



Prediction 1: Accuracy and Time. Inference



Not making general claim: specific instances in experiment.
Details: Informativeness and Signal Strength.

Accuracy and Time

Nonmonotonicity because Participants Quitting in Complex Conditions?

No, they still exert effort, just less. Complex conditions not fastest. (Details)

Also: patterns for RT robust to outliers (e.g., median instead of mean).

Fast or Slow Errors? (Accuracy | RT) increasing/decreasing in RT, depending on task.

(More)

Prediction 1': Accuracy is monotone in ability; time is single-peaked.

Time has single-crossing property in ability and problem complexity.

Proxy for ability for task with participant average accuracy (symmetric subsidies).

Single-crossing: Slower more predictive of higher ability in more complex problems than in simpler ones.

Prediction 2: Effect of Subsidies on Choices and Time

Prediction 2: Effect of Subsidies on Choices and Time

Subsidising an alternative leads to it being chosen more often and faster, and to choosing other alternatives less often and slower.

Testing Prediction:

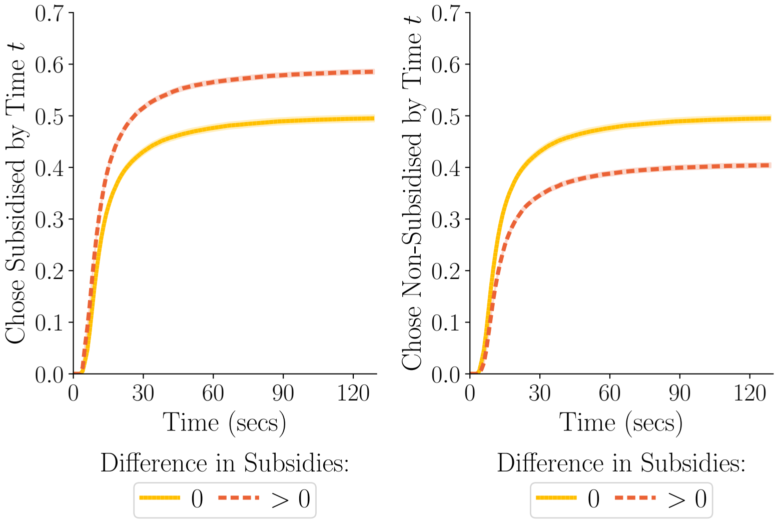
Nonparametric: Frequency Chose Alternative before t by Subsidy Difference.

Also Avg Effects:

(1) Chose Subsidised by Subsidy Difference.

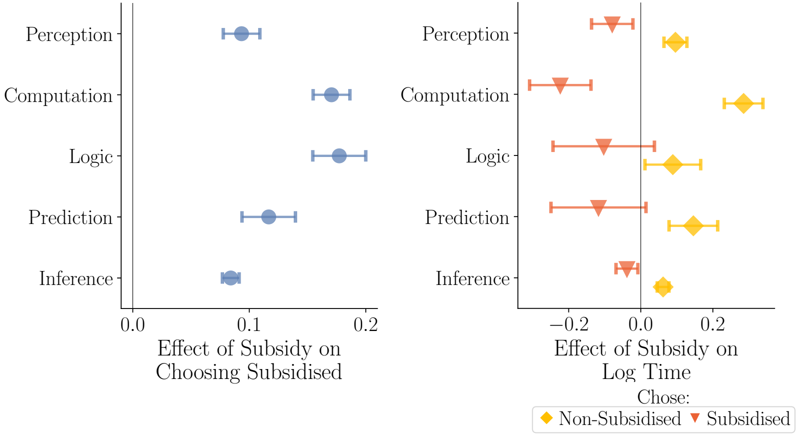
(2) Log(RT) by Subsidy Difference.

Prediction 2: Effect of Subsidies on Choices and Time



Pooled data. Disaggregated data all consistent: (Perception), (Computation), (Logic), (Prediction), (Inference). Pointwise dominance of cumulative incidence functions.

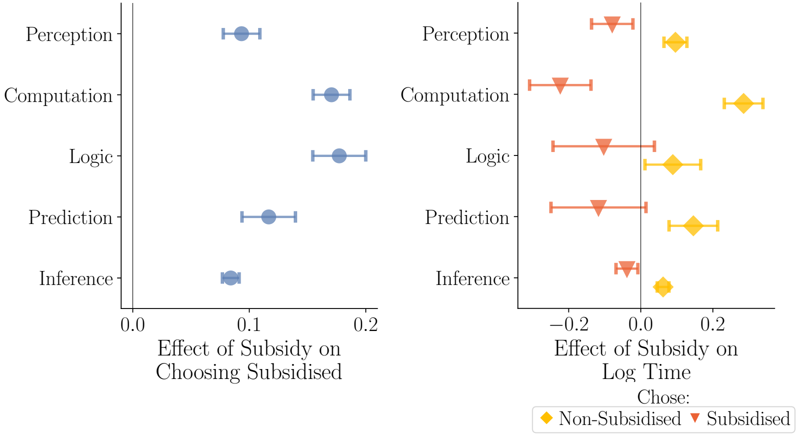
Prediction 2: Effect of Subsidies on Choices and Time



Chose Subsidised = $\beta_0 + \beta_1$ Subsidy

Log RT = $\beta_0 + \beta_1$ Subsidy + β_2 Chose Subsidised $\times$ Subsidy

Prediction 2: Effect of Subsidies on Choices and Time



Subsidising an alternative leads to it being chosen more often and faster, and to choosing other alternatives less often and slower.

Prediction 3: Revealing Complexity

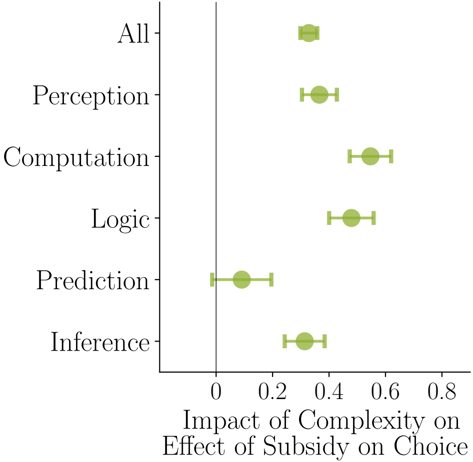
Prediction 3: Revealing Complexity

Choices are more responsive to subsidies in more complex problems.

Testing Prediction:

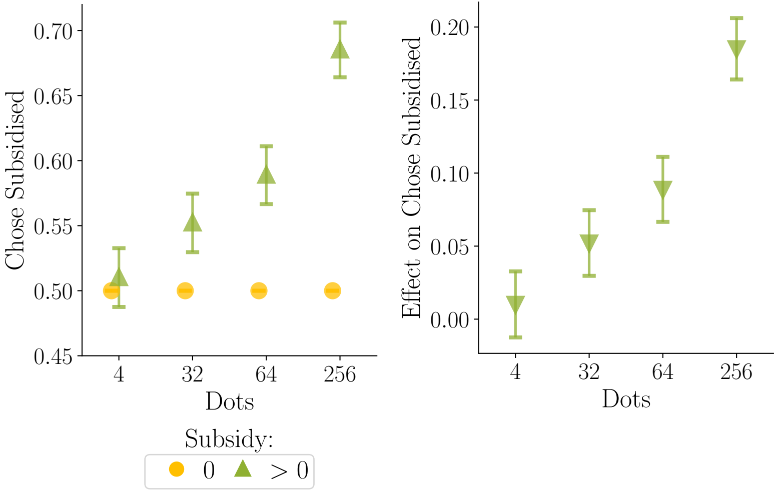
- Operationalising complexity: Task Complexity = 1-Task Avg Accuracy for all conditions (with equal subsidies).
- Regression: Chose Subsidised = $\beta_0 + \beta_1$ Task Complexity (with different subsidies).
- Comparing means, by condition and subsidy.
- How much more chose alternative as function of subsidy difference.

Prediction 3: Revealing Complexity



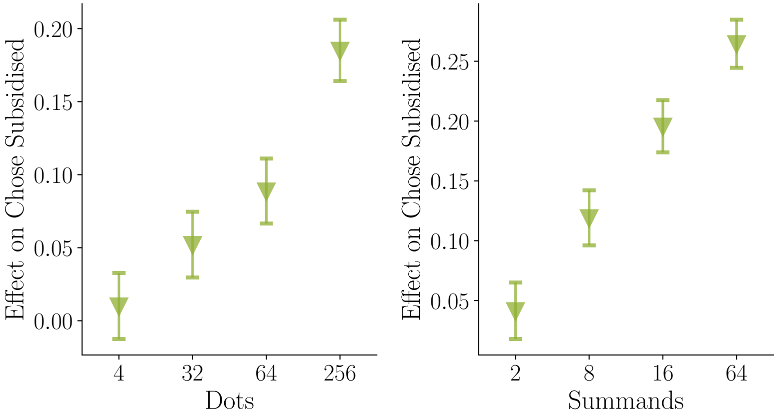
Regression: Chose Subsidised = $\beta_0 + \beta_1$ Task Complexity.

Prediction 3: Revealing Complexity. Disaggregated



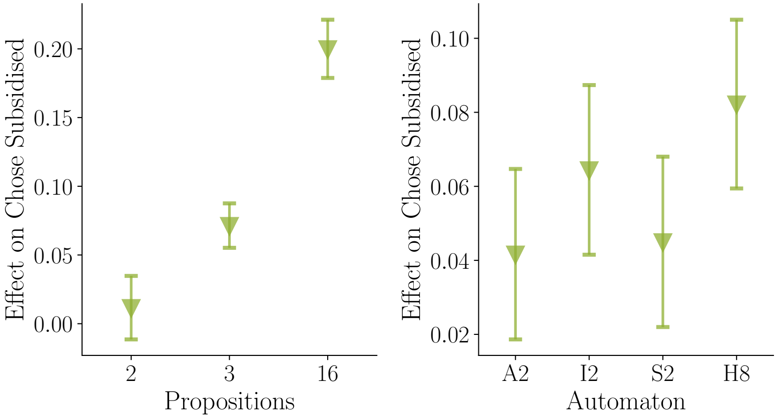
No Subsidy → Arbitrary label, 'Choice Subsidised'=.5. Normalise to 0.

Prediction 3: Revealing Complexity. Perception and Computation



Choices are more responsive to subsidies in more complex problems.

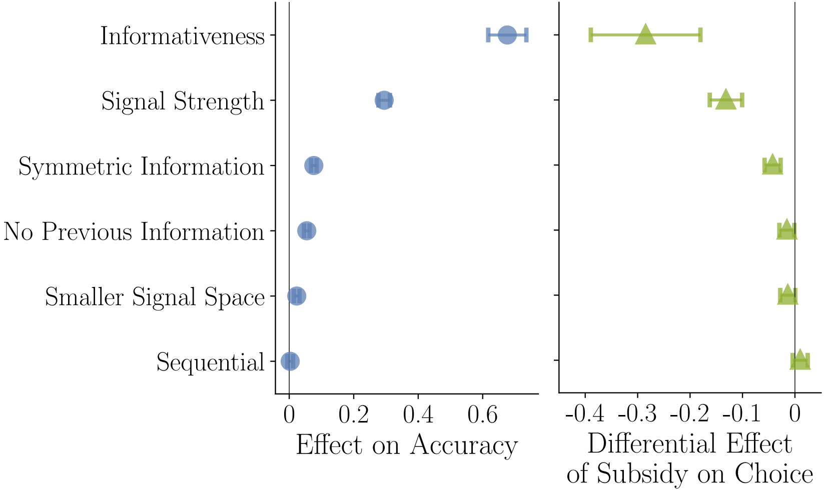
Prediction 3: Revealing Complexity. Logic and Prediction



Choices are more responsive to subsidies in more complex problems.

Meta-cognition crucial. More on this later.

Prediction 3: Revealing Complexity. Inference



Choices are more responsive to subsidies in more complex problems.

Details: Informativeness and Signal Strength.

Prediction 4: Revealing Ability

Prediction 4: Revealing Ability

Less able DMs' choices are more responsive to subsidies.

Testing Prediction:

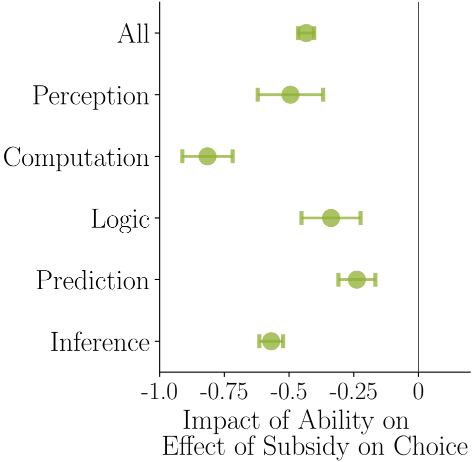
Operationalising ability: Ability = Individual Avg. Accuracy for whole experiment (with equal subsidies).

ECDF Avg Accuracy: (Perception), (Computation), (Inference), (Logic), (Prediction).

Regression: Chose Subsidised = $\beta_0 + \beta_1$ Individual Ability (with different subsidies).

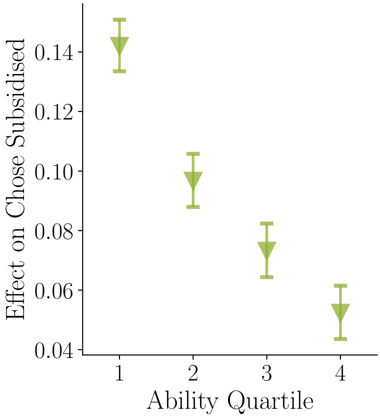
Comparing means, by condition and subsidy.

Prediction 4: Revealing Ability



Regression: Chose Subsidised = $\beta_0 + \beta_1$ Individual Ability.

Prediction 4: Revealing Ability



Less able participants' choices are more responsive to subsidies.

Pooled data. Disaggregated data all consistent: (Perception), (Computation), (Logic), (Prediction), (Inference).

Meta-Cognition and Learning

Meta-Cognition and Learning

Meta-cognition crucial:

In Prediction task, hard to distinguish A2, I2, and S2.

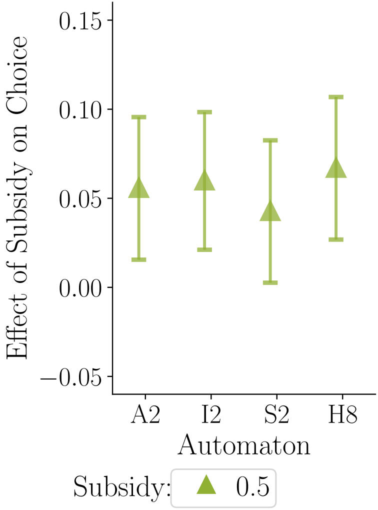
No indication as in Perception and Computation.

No visual difference in urns as in Inference.

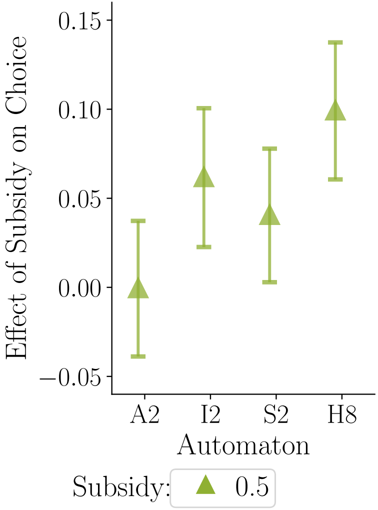
No difference in text length as in Logic.

Nevertheless: evidence that participants learn to identify complexity...

Learning to Identify Complexity



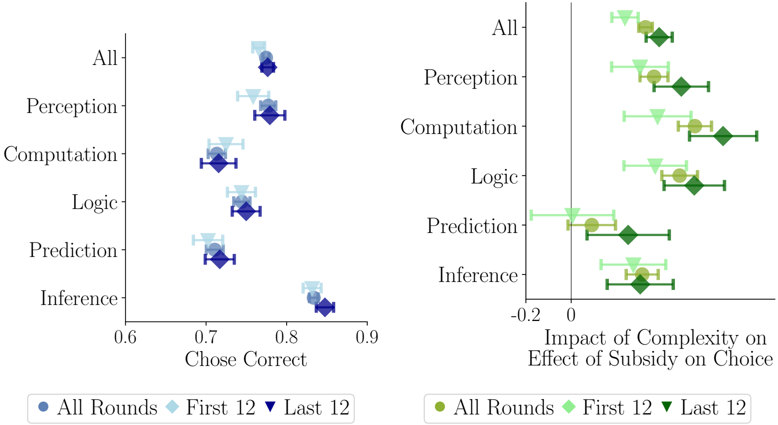
1st 12 Rounds



Last 12 Rounds

Evidence that participants learn to identify complexity.

Learning to Identify Complexity



Evidence that participants learn to identify complexity (react more to subsidies when task is more complex).

But no significant improvement in performance over time.

Also no significant effect on meta-cognition about own ability. (Details)

Prediction 0: Exogenous Stopping

Prediction 0: Exogenous Stopping

Endogenous Stopping: feature affecting problem complexity independent from which alternative is optimal.

Accuracy decreases in problem complexity, time non-monotone.

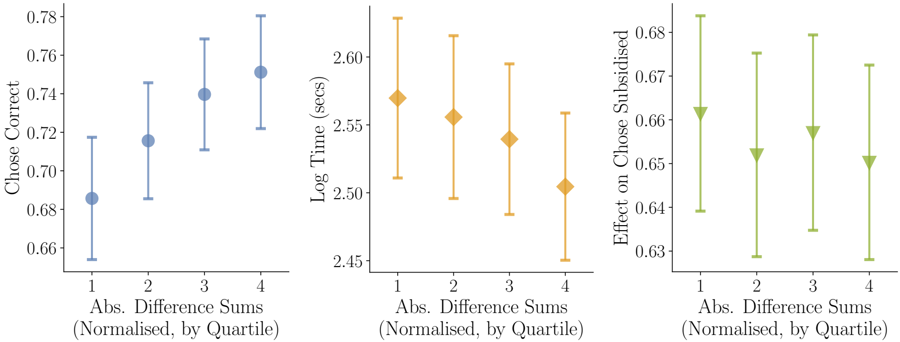
E.g., sums with more/fewer summands.

Exogenous Stopping: feature affecting problem complexity also determines which alternative is optimal.

Accuracy decreases in problem complexity, time increases in problem complexity.

E.g., absolute difference in sums.

Prediction 0: Exogenous Stopping



Limits of method: need to recognise and adjust to problem complexity.

Related Literature

Related Literature

Experiments on Complexity: Hogarth 75; Wilcox 93; Rubinstein 07, 08, 13; Bossaerts Murawski 17; Caplin Csaba et al. 20; Oprea 20, 24; Kendall Oprea 24; **Enke Graeber Oprea WP**, 25; Esteban-Casanelles Gonçalves WP; Gonçalves Libgober Willis 25; **Agranov Schotter Trevino WP**; Agranov Reshidi WP; Arrieta Nielsen WP; Puri 25; Shubatt Yang WP; Enke Shubatt WP; Musolff Zimmermann WP; . . .

Sequential Sampling: Wald 45; Dvoretzky Kiefer Wolfowitz 53; Peskir Shiryaev 06; Moscarini Smith 01; Drugowitsch et al. 12; Halac Kartik Liu 16, 17; Bobtcheff Levy 17; Steiner Stewart Matějka 17; Morris Strack WP; Fudenberg Strack Strzalecki 18; **Gonçalves WP**; . . .

DDM: Ratcliff 78; Schouten Bekker 67; Wickelgren 77; Bogacz et al. 06; Fehr Rangel 11; Krajbich Lu Camerer Rangel 12; Caplin Martin 11; Ratcliff Smith Brown McKoon 16; Forstmann Ratcliff Wagenmakers 16; Bhui 19; Webb 18; Alós-Ferrer Fehr Netzer 22; . . .

This paper: Method to infer problem complexity and agent ability from choices alone.

Conclusion

Conclusions

This paper: Experimentally test a method to infer problem complexity and agent ability from choices alone.

- **Portable measure:** how much choices are affected by subsidies increases in problem complexity and decreases in DMs' ability.
- **Applied to wide range of problem domains:** perception, computation, deduction, prediction, inference.
- Based on sequential sampling framework.
- **Predictions involving choices and response time supported by the data.**
- Response time not good proxy for complexity.
- Importance of meta-cognition and learning (a problem's) complexity.

Also in the paper: more detailed analyses, robustness checks (comprehension checks, removing time outliers as per pre-registrations).

Revealing Complexity

Duarte Gonçalves
University College London

Julen Zarate-Pina
University of the Basque Country

Salvatore Nunnari
Bocconi University

Behavioral/Experimental Seminar

University of Pittsburgh

17 April 2026

Overview

1. Why Not Just Use Response Times?
2. A Method to Infer Complexity (And Ability)
3. Experimental Design
   - Task Domains
   - Perception
   - Computation
   - Logic
   - Prediction
   - Other Experimental Details
   - Inference
4. Prediction 1: Accuracy and Time
5. Prediction 2: Effect of Subsidies on Choices and Time
6. Prediction 3: Revealing Complexity
7. Prediction 4: Revealing Ability
8. Meta-Cognition and Learning
9. Prediction 0: Exogenous Stopping
10. Related Literature
11. Conclusion

Inference

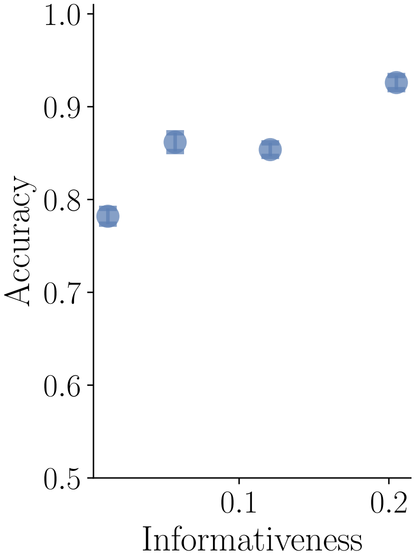
Balls-and-urns paradigm (Edwards & Phillips 64; cf. Benjamin 19). (Back)

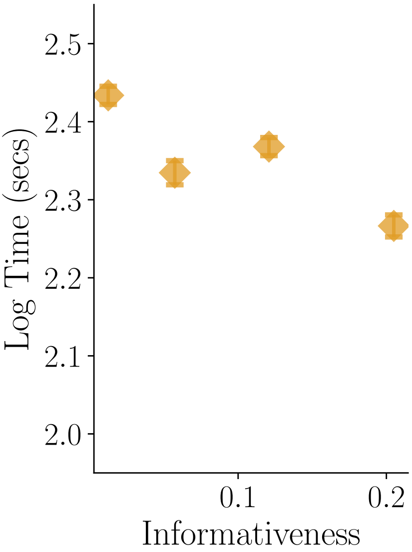
- Two base urns compositions:
  A = (24,76) and B = (36,64) + symmetric: A' = (76,24) and B' = (64,36).
- All pairwise combinations (up to relabelling): (AA'), (AB), (AB'), (BB'). Uniform prior.
- Pairs of urns with 4 signals/colours: split each signal into 2 different signals with same likelihood.

E: urn 1/2 is more likely.

Prediction 1: Accuracy and Time. Inference: Informativeness

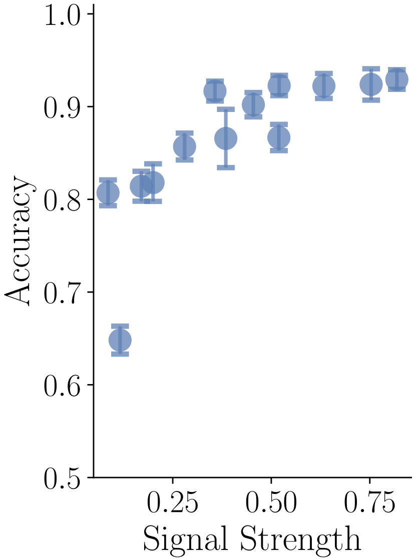
(Back)

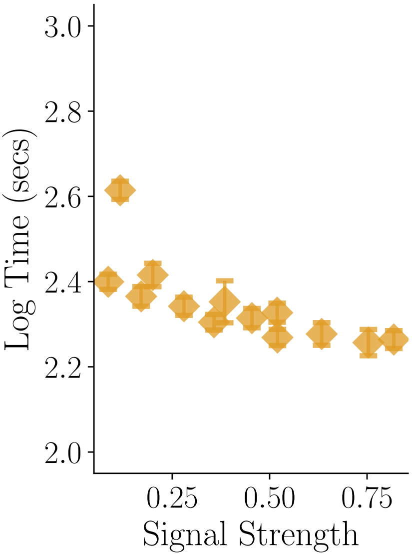




Prediction 1: Accuracy and Time. Inference: Signal Strength

(Back)





Quitting or Reducing Effort?

(Back)

Quitting when most complex $\implies$ Fastest RT when most complex.

If N rounds of most complex: N fastest rounds correspond to most complex condition.

No one consistently quits with most complex conditions.

Quitting or Reducing Effort?

(Back)

Quitting when most complex $\implies$ Fastest RT when most complex.

If N rounds of most complex: N fastest rounds correspond to most complex condition.

No one consistently quits with most complex conditions.

Weaker condition: 75% of N fastest rounds correspond to most complex condition.

Perception: <1%. Computation: 1.4%. Logic: 3.7%. Prediction: <1%.

Very weak condition: 2/3 of N fastest rounds correspond to most complex condition.

Perception: 5.1%. Computation: 5.1%. Logic: 10%. Prediction: <1%.

Accuracy and Time

Nonmonotonicity because Participants Quitting in Complex Conditions?

No, they still exert effort, just less. Complex conditions not fastest. (Details)

Also: patterns for RT robust to outliers (e.g., median instead of mean).

Fast or Slow Errors? (Back)

(Accuracy | RT) increasing and decreasing in RT, depending on task.

- Perception and Computation: slower is better.
- Inference: faster is better.
- Logic and Prediction: mixed evidence.

Prediction 1': Accuracy is monotone in ability; time is single-peaked.

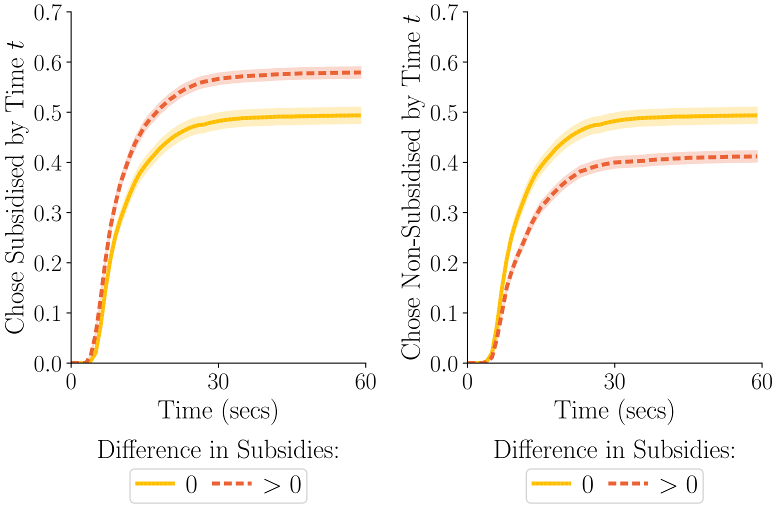
Time has single-crossing property in ability and problem complexity.

Proxy for ability for task with participant average accuracy (symmetric subsidies).

Single-crossing: Slower more predictive of higher ability in more complex problems than in simpler ones.

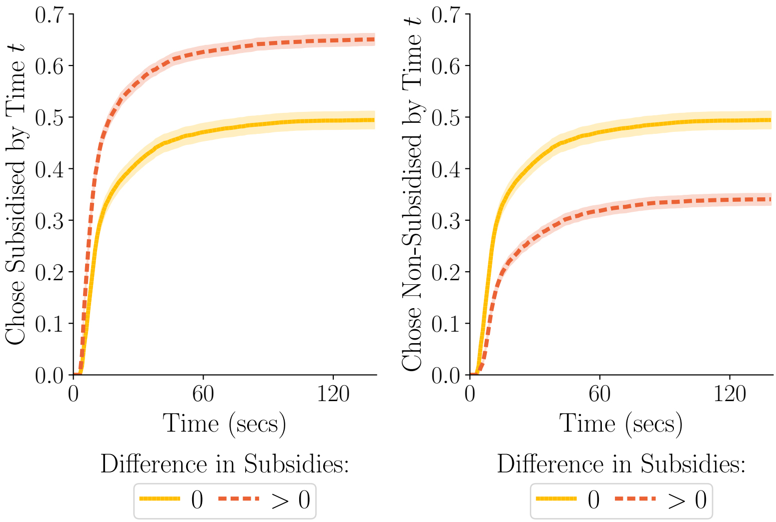
Prediction 2: Effect of Subsidies on Choices and Time. Perception

(Back)



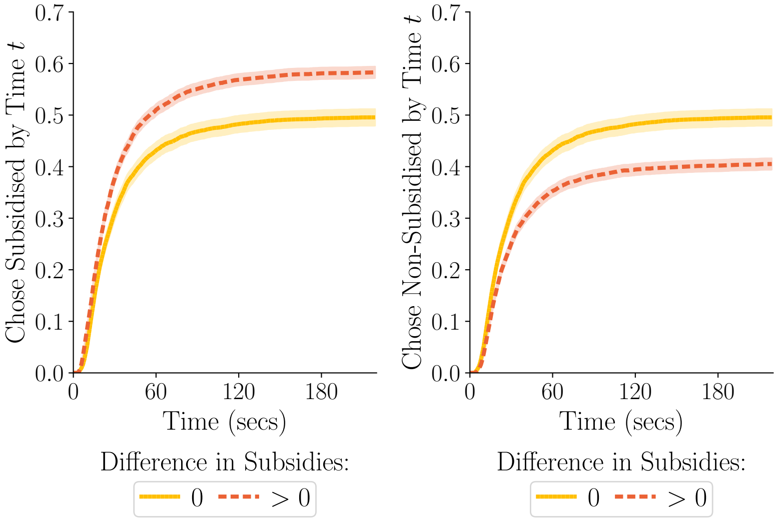
Prediction 2: Effect of Subsidies on Choices and Time. Computation

(Back)



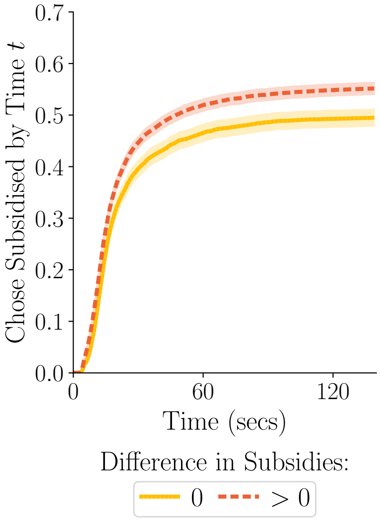
Prediction 2: Effect of Subsidies on Choices and Time. Logic

(Back)



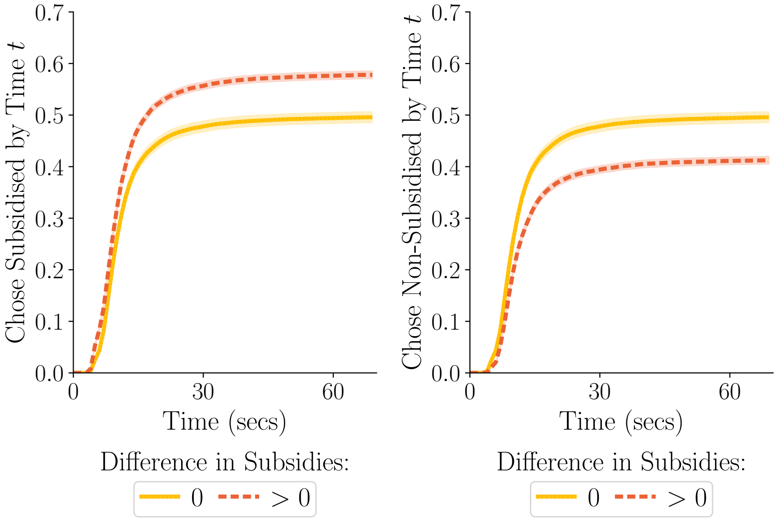
Prediction 2: Effect of Subsidies on Choices and Time. Prediction

(Back)

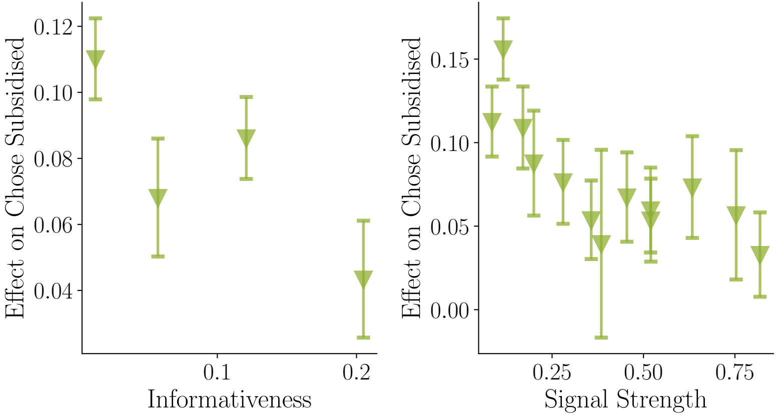


Prediction 2: Effect of Subsidies on Choices and Time. Inference

(Back)



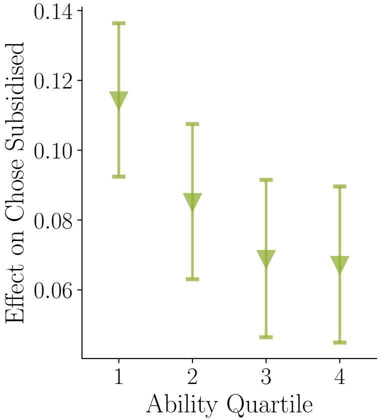
Prediction 3: Revealing Complexity. Inference: Informativeness and Signal Strength (Back)



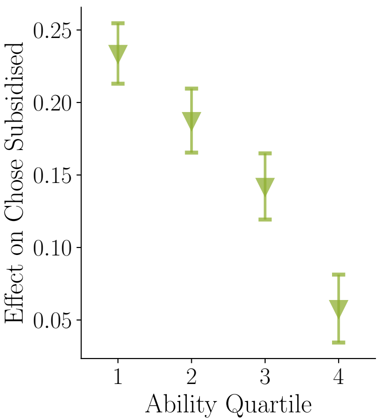
Choices are more responsive to subsidies in more complex problems.

Prediction 4: Revealing Ability. Perception and Computation

(Back)



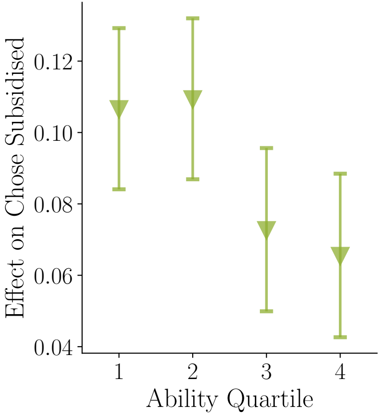
Perception



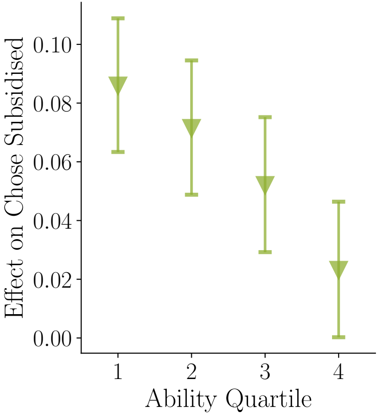
Computation

Prediction 4: Revealing Ability. Logic and Prediction

(Back)



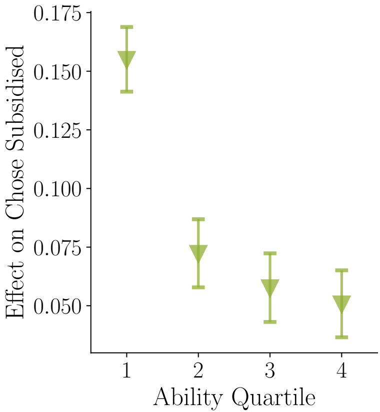
Logic



Prediction

Prediction 4: Revealing Ability. Inference

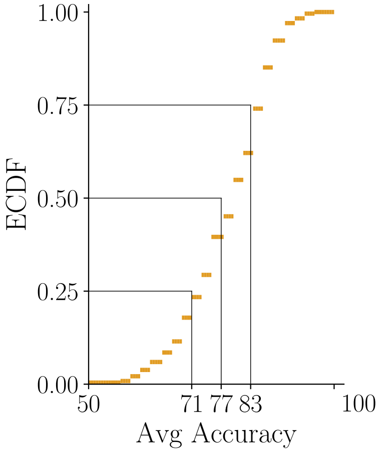
(Back)



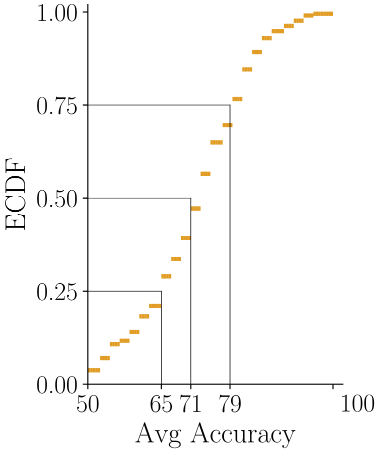
Inference

Ability and Avg Accuracy: Perception and Computation

(Back)



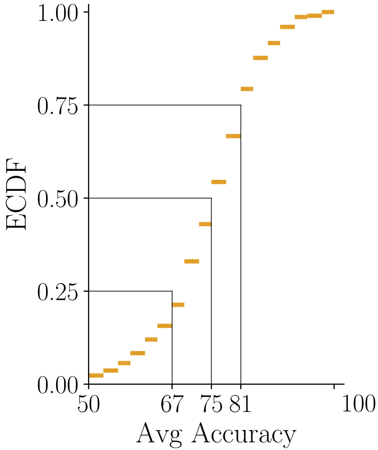
Perception



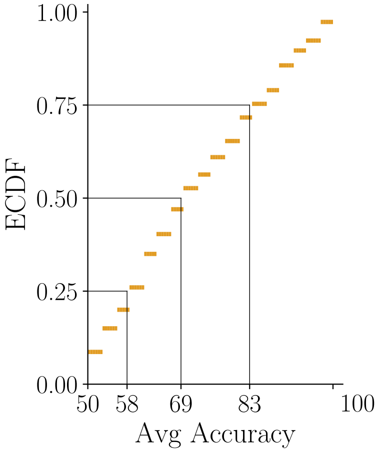
Computation

Ability and Avg Accuracy: Logic and Prediction

(Back)



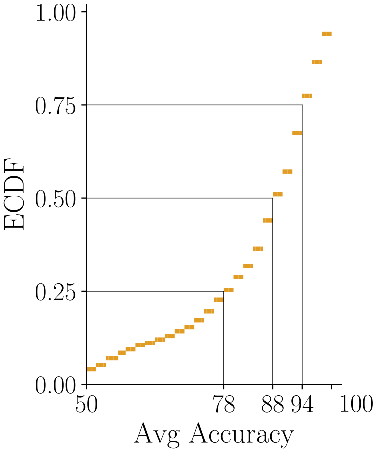
Logic



Prediction

Ability and Avg Accuracy: Inference

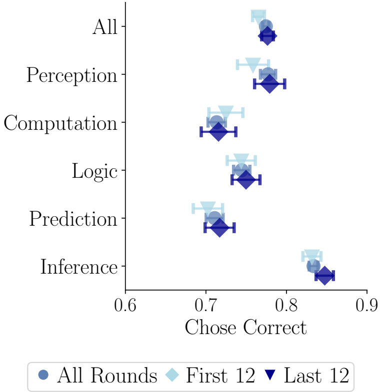
(Back)

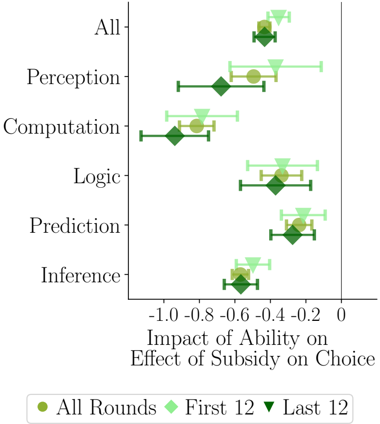


Inference

Learning Own Ability

(Back)





No evidence that participants learn own ability.