Time Costs and Behaviour Across Participant Pools

Duarte Gonçalves[*] ⓡ Terri Kneeland[†] ⓡ Julen Zarate-Pina[‡]

Abstract

University laboratory samples often produce cleaner experimental data than broader online samples such as Prolific. We examine whether this gap reflects differences in the opportunity cost of time. In an experiment spanning laboratory and online participant pools, we exogenously vary both the cost of deliberation and task complexity. Online participants are faster and less accurate at baseline. Raising time costs in the laboratory lowers accuracy, while reducing the return to answering quickly online improves it. Participant-pool differences arise where deliberation is both costly and productive, linking external validity to experimental design and task complexity.

Keywords: Participant Pools; Online Experiments; Response Time; Costly Cognition; Task Complexity.
JEL Classifications: C83, D83, D90, D91.

[*] Department of Economics, University College London; duarte.goncalves@ucl.ac.uk.
[†] Department of Economics, University of Calgary; t.kneeland@ucalgary.ca.
[‡] Department of Economic Analysis, University of the Basque Country; julen.zarate-pina@ehu.eus.
The project was supported by ERC Grant SUExp - 801635 and by the Basque Government under Grant IT1697-22.
ⓡ denotes randomised authorship order (Ray ⓡ Robson, 2018).
This draft: 12 September 2026.

1. Introduction

Experimental results often differ across participant pools. A growing literature finds that participants recruited through online platforms generate noisier data than university students participating in laboratory experiments (Snowberg and Yariv, 2021; Gupta, Rigotti, and Wilson, 2021). These differences are commonly attributed to participant composition: laboratory and online participants may differ in preferences, beliefs, abilities, or decision rules. This interpretation underlies both concerns about the external validity of laboratory experiments and researchers' decisions about which populations to recruit. Yet changing participant pools generally changes more than who participates: it also changes the environment in which participants make decisions.

One often neglected difference between laboratory and online *environments* is the opportunity cost of deliberation. Laboratory participants typically attend approximately fixed-duration sessions and cannot leave immediately after completing a study. Online participants can instead proceed to another paid study as soon as they finish, so the marginal cost of deliberating longer may be higher online.

When accurate choices require deliberation, differences in time costs can induce systematic differences in behaviour across participant pools. A participant can deliberate for longer about which action is best before choosing. Higher time costs reduce the net value of continued deliberation and may therefore lead to both faster and less accurate choices. This mechanism is formalised in models of optimal stopping and sequential sampling (Fehr and Rangel, 2011; Fudenberg et al., 2018; Gonçalves, 2026). These models have proved useful for rationalising empirical relationships between choices and response times in economics (Clithero, 2018; Schotter and Trevino, 2021; Alós-Ferrer et al., 2021; Chiong et al., 2023), as well as in cognitive science and psychology (Forstmann et al., 2016; Krajbich et al., 2026). Participant-pool comparisons may therefore reflect both participant composition and time costs.

The behavioural consequences of time costs should depend on task complexity.[1] In very simple problems, the payoff-maximising action becomes apparent quickly; in extremely difficult problems, additional deliberation may do little to improve choices, so participants may also stop quickly. This non-monotonic relationship between complexity and deliberation time has recently been characterised theoretically (Gonçalves, 2026) and documented empirically across several tasks (Gonçalves et al., 2026; Agranov et al., 2025). Differences in time costs should matter most between these extremes, where additional deliberation is productive. The mechanism therefore generates a distinctive prediction: differences in choice quality across participant pools and time-cost environments should be concentrated at intermediate levels of task complexity. Accordingly, null differences across participant pools in very easy or difficult tasks need not extend to where deliberation has a higher return.

[1] See Oprea (2025) for a comprehensive discussion of task complexity.

This paper tests this mechanism in an experiment with university students recruited for laboratory sessions and online participants recruited through Prolific. Participants complete two incentivised tasks with objectively correct answers, allowing us to measure choice quality directly, while task complexity varies within participant. We first show that online participants are less accurate than laboratory participants, reproducing a standard baseline pattern, and, in line with our mechanism, they also respond more quickly. We then change the time-cost environment while holding the participant pool fixed. Discounting laboratory earnings based on response time causes laboratory participants to choose faster and less accurately. Imposing a delay on online participants' choices increases their accuracy. In both cases, the effects on accuracy are concentrated at intermediate levels of complexity, where deliberation is most valuable.

The experiment comprises two baseline conditions comparing laboratory and online environments and two exogenous time-cost manipulations within each pool. In the laboratory, we increase the cost of deliberation by discounting the payoff from a correct choice as a function of response time. Online, we reduce the opportunity cost of deliberation by imposing a page-progression delay. This design allows us to document naturally-occurring differences across participant pools and test whether time-cost changes move behaviour as predicted. The tasks, adapted from Gonçalves et al. (2026), vary perceptual and computational complexity. The within-subject variation in complexity allows us to test how the effects of time costs depend on the value of additional deliberation.

We report three main findings. First, in the baseline treatments, online participants respond faster and choose less accurately than laboratory participants. Laboratory participants achieve average choice accuracy of 86%—the frequency with which they choose the payoff-maximising alternative—with an average response time of 34.5 seconds. In contrast, online participants are 9 percentage points (p.p.) less likely to choose the best alternative and take between 25-40% less time. These results reproduce the benchmark finding of noisier data from participants on online platforms and show that lower accuracy is accompanied by shorter response times.

Second, the experimental manipulations move behaviour in the direction predicted by the time-cost mechanism. Increasing the cost of time in the laboratory significantly reduces response times by 68% and lowers accuracy by 13 p.p. Imposing a delay online induces participants to take longer to respond and moves behaviour in the opposite direction: accuracy increases by 5 p.p. Thus, decision quality changes when the timing environment changes, even when the participant pool is held fixed.

Third, participant-pool and treatment differences depend on task complexity. The mechanism predicts the largest differences at intermediate complexity, where additional deliberation is most likely to improve accuracy. The data are consistent with this prediction. We find little evidence of baseline participant-pool differences in either very easy or very hard problems. In very easy problems, response times are fast and accuracy is very high and virtually indistinguishable across participant pools and time-cost manipulations. In very hard problems, responses are again short and accuracy is close to uniform across pools and time-cost manipulations. The differences are instead concen-

trated at intermediate levels of complexity, precisely where lower time costs should induce more deliberation and improve choice quality.

Taken together, the results show that noisier online data are not simply an inherent property of online participants. Decision quality responds to the incentives and frictions governing deliberation. Participant-pool comparisons therefore compare both recruited populations and environments with different opportunity costs of time. External validity depends on whether the participant pool, task complexity, and time-cost environment jointly approximate the target economic setting. In some settings, high time costs and rapid responses may be externally valid; in others, greater scope for deliberation may provide the more relevant benchmark.

The findings also have a practical design implication. A page-progression delay can be implemented in experiments and surveys, and in our setting it increases choice accuracy. Comparing behaviour with and without a delay may provide a diagnostic test of whether incentives for rapid completion contribute to decision noise.

Finally, the results have a measurement implication: objective measures of decision quality are task-specific. Researchers sometimes include choices with objectively correct answers—such as strictly dominated strategies or lotteries—to diagnose choice quality in experiments whose main outcomes have no objectively correct answer. Such diagnostics are informative only to the extent that their complexity resembles that of the substantive choices. A simple dominance check may understate noise in more complex decisions, whereas an excessively difficult diagnostic may overstate noise in simpler ones.

1.1. Related Literature

Our paper contributes to three literatures.

First, we contribute to work comparing behaviour and data quality across participant pools. Early online experiments often reproduced qualitative laboratory findings (Paolacci et al., 2010; Horton et al., 2011), and comparative statics frequently generalise even when point estimates do not (Fréchette, 2017). More recent evidence documents meaningful differences across laboratory, MTurk, Prolific, and broader population samples (Snowberg and Yariv, 2021; Gupta et al., 2021). These differences are usually interpreted as consequences of participant composition, including differences in preferences, experience, cognitive ability, or decision rules. However, lower online choice quality need not primarily reflect inattention or poor compliance (Çelebi et al., 2026). This distinction matters because heterogeneous choice noise can generate spurious relationships between participant characteristics and elicited preferences (Andersson et al., 2016). We show experimentally that behaviour attributed to a participant pool responds to deliberation costs even when the pool is held fixed.

Second, our paper relates to research on time pressure, response time, and deliberation; see Spiliopoulos and Ortmann (2018) for a review. Experimental manipulations of decision time affect behaviour

in social, risky, and strategic choices, although their direction varies across domains (Kocher and Sutter, 2006; Kocher et al., 2013; Kirchler et al., 2017; Haji et al., 2019; Fromell et al., 2020; Imas et al., 2022). Particularly relevant for our interpretation, Baillon et al.'s (2018) finding that time pressure increases ambiguity insensitivity, rather than ambiguity aversion, suggests time costs affect the precision of information processing rather than underlying preferences.

In contrast, we study tasks with objectively correct choices, allowing treatment effects to be interpreted as changes in decision quality rather than measured preferences. Our interpretation builds on sequential-sampling and optimal-stopping models in which participants trade off the expected improvement from further deliberation against its cost (Fehr and Rangel, 2011; Fudenberg et al., 2018; Gonçalves, 2026). Whereas much of the response-time literature uses decision time to recover preference intensity, resolve, or latent choice processes (Clithero, 2018; Alós-Ferrer et al., 2021; Schotter and Trevino, 2021; Liu and Netzer, 2023), we emphasise that response times also reflect deliberation's opportunity cost, which may vary across participant pools and elicitation environments.

Third, we contribute to the emerging literature on complexity in economic choice (Oprea, 2025). This literature views complexity as the resource burden of implementing the procedures required to solve a task and studies how people respond by using simpler procedures, making mistakes, or avoiding cognitively demanding decisions (Oprea, 2020; Bossaerts and Murawski, 2017; Enke and Graeber, 2023; Sanjurjo, 2025; Enke et al., 2026). An important implication is that response time need not increase monotonically with complexity: participants may deliberate extensively on intermediate tasks but less so when problems become very easy or extremely difficult (Gonçalves, 2026; Gonçalves et al., 2026; Agranov et al., 2025). Our contribution is to connect complexity to the opportunity cost of time. We show that time-cost and participant-pool differences are concentrated where additional deliberation is productive, rather than being uniform across tasks.

2. Conceptual Framework and Hypotheses

We develop a reduced-form model of decision-making in experiments. The model isolates a difference between laboratory and online participants that is central to our analysis: the opportunity cost of deliberation time.

2.1. Choices and Deliberation Time

In experiments, participants may expend effort to identify the action that maximises their experimental payoff. Let $a \in [0,1]$ denote choice accuracy, the probability of selecting the payoff-maximising action or the extent to which higher-payoff actions are chosen.[2] We express the experiment's incentivisation scheme as $U : [0,1] \to \mathbb{R}_+$, so that $U(a)$ denotes the payoff that the participant receives given choice quality a.

Choice accuracy depends on the time spent deliberating and on task complexity. Let $T \in [0, \overline{T}]$ de-

[2]The same reduced-form formulation can be used to represent choice noise or costly attention.

note deliberation time, where $\overline{T} > 0$ is the maximum feasible duration, and let $\kappa > 0$ denote task complexity. Following sequential-sampling models that relate decision time to choice quality (see, e.g., Fudenberg et al., 2018; Gonçalves, 2026), we write $a = A(T/\kappa)$. Thus, longer deliberation improves accuracy, whereas greater complexity reduces effective deliberation for a given amount of time. We assume that A and U are twice continuously differentiable, strictly increasing, and strictly concave. To rule out a zero-deliberation corner, we also impose the Inada-type condition $A'(0^+) = \infty$.

Time enters the model through costly effort and delay. First, longer deliberation improves outcomes but is costlier, reflecting a speed-accuracy trade-off. We define $C : \mathbb{R}_+ \to \mathbb{R}_+$ as the (discounted) cost to total effort, satisfying $C \in \mathscr{C}^2$, $C(0) = 0$, $C' > 0$, and $C'' > 0$. Second, spending longer in the experiment delays continuation activity, such as leisure, work, or another paid experiment. Let $V > 0$ denote the value of that activity, and $r > 0$ the participant's discount rate. We assume that $r\overline{T} \leq 1$.

2.2. Time Costs in Laboratory and Online Experiments

In laboratory experiments, an individual participant's deliberation time may have little effect on when they can leave. Laboratory sessions often involve several participants, payments are typically made after the session, and participants remain seated until others finish. We therefore model the end of the session as occurring at the approximately fixed time $\overline{T}$.[3] A laboratory participant solves

$$\max_{T \in [0,\overline{T}]} \exp(-r\overline{T})U(A(T/\kappa)) - C(T) + \exp(-r\overline{T})V. \tag{Lab Baseline}$$

The participant chooses T by trading off the return to greater accuracy against the cost of deliberation. The continuation value is discounted to the end of the session and therefore does not vary with the participant's own deliberation time.

For participants recruited through online platforms, the opportunity cost of deliberation is higher than for laboratory participants. Although payment for the current study may be delayed, a participant can typically begin another activity as soon as the study is completed. Additional deliberation therefore postpones the continuation value directly, inducing an explicit opportunity cost associated with making slower choices. We represent the online participant's problem as

$$\max_{T \in [0,\overline{T}]} \exp(-r\overline{T})U(A(T/\kappa)) - C(T) + \exp(-rT)V. \tag{Online Baseline}$$

The two environments may differ along many other dimensions, including participant composition. The model isolates the time-cost wedge that arises even in the absence of such differences. For an interior solution, the first-order condition for the problem faced by online participants is

$$\exp(-r\overline{T})U'(A(T^*/\kappa))\frac{A'(T^*/\kappa)}{\kappa} - C'(T^*) - r\exp(-rT^*)V = 0,$$

where the terms are, respectively, the marginal experimental benefit, effort cost, and opportunity

[3] The maximum session duration $\overline{T}$ need not bind in every session. The relevant assumption is that an individual's deliberation time has limited effect on when they leave the laboratory. In practice, the session duration may depend on the slowest participant. We abstract from the resulting strategic interaction.

cost of delaying continuation. The laboratory first-order condition is identical except that it omits the final opportunity-cost term, since the participant's own decision time does not affect the end of the session. The corresponding first-order condition in the laboratory is

$$\exp(-r\overline{T})U'(A(T^*/\kappa))\frac{A'(T^*/\kappa)}{\kappa} - C'(T^*) = 0,$$

where the marginal opportunity-cost term is absent because the participant's own decision time does not affect the end of the session.

Under the stated assumptions, the online participant faces the same marginal benefit of deliberation as the laboratory participant, but an additional marginal cost from delaying the continuation activity. This additional cost lowers the optimal amount of deliberation relative to the laboratory environment. Since choice accuracy A is increasing in deliberation time, the reduction in deliberation also lowers choice accuracy.

This yields our first hypothesis.

Hypothesis 1. Decision time and choice accuracy are lower in online experiments than in laboratory experiments.

Existing evidence indicates that choices made by online participants are noisier than those made by laboratory participants (Snowberg and Yariv, 2021). Hypothesis 1 identifies one mechanism that can generate this pattern: costly deliberation combined with a higher opportunity cost of time.

2.3. Manipulating Time Costs

The model suggests two interventions that, by altering time costs while holding the participant pool fixed, can level the field without requiring the estimation of idiosyncratic preferences.

First, the experimentalist can increase the time cost of deliberation in the laboratory by discounting the payoffs accrued in the experiment at rate $\rho > 0$. The laboratory participant then solves

$$\max_{T\in[0,\overline{T}]} \exp(-r\overline{T} - \rho T)U(A(T/\kappa)) - C(T) + \exp(-r\overline{T})V. \qquad \text{(Lab Discounting)}$$

Relative to Lab Baseline, discounting lowers the marginal return to continued deliberation. It therefore reduces optimal deliberation time and, since A is increasing, also reduces choice accuracy.[4]

Second, the experimentalist can impose a forced delay, a minimum response time $\underline{T} > 0$ for online participants. A participant who finishes deliberating before $\underline{T}$ cannot begin the continuation activity until the delay has elapsed. The participant solves

$$\max_{T\in[0,\overline{T}]} \exp(-r\overline{T})U(A(T/\kappa)) - C(T) + \begin{cases} \exp(-r\underline{T})V, & T \le \underline{T}, \\ \exp(-rT)V, & T > \underline{T}. \end{cases} \qquad \text{(Online Delay)}$$

[4]The treatment discounts the payoff generated by the participant's choice rather than a fixed lump-sum payment; this is deliberate. This distinction may be relevant when ability is positively correlated with the value of outside opportunities. Discounting the choice-contingent payoff in the laboratory makes the marginal cost of deliberation depend on the value generated by the choice and may therefore more closely reproduce this pattern.

For $T < \underline{T}$, additional deliberation does not further delay the continuation activity. The marginal opportunity cost of time is therefore absent over this range, as in the laboratory problem. Participants who would otherwise stop before $\underline{T}$ now have an additional interval during which deliberation does not delay their continuation activity. The delay consequently weakly increases optimal deliberation time. Since A is increasing, choice accuracy weakly increases. This does not require participants to spend the forced waiting time deliberating; rather, it removes the return to submitting immediately.

This yields our second hypothesis.

Hypothesis 2. Discounting payoffs reduces decision time and choice accuracy, whereas imposing a minimum response time increases both.

Whether the laboratory discounting treatment quantitatively reproduces the online baseline, or the online delay treatment reproduces the laboratory baseline, depends on the values of ρ and $\underline{T}$. The hypotheses concern the direction of the treatment effects rather than exact equivalence across environments.

2.4. Time Costs and Task Complexity

The value of deliberation depends on task complexity. In sequential-sampling models, greater complexity generally lowers accuracy, but optimal deliberation time need not increase monotonically with complexity (cf. Gonçalves, 2026). Very simple tasks require little deliberation; in very complex tasks, additional deliberation may have little effect on accuracy. Deliberation time may therefore be greatest at intermediate complexity, where additional time produces the largest improvement in expected choice quality. In line with this class of models, here too it can be easily shown that choice accuracy $A(T^*/\kappa)$ decreases in task complexity κ, but optimal deliberation time T^* is in fact non-monotone in task complexity, being lowest for very simple and very complex tasks. This non-monotonic relationship has been documented by Gonçalves et al. (2026) across several tasks, including those used here.

The same logic has meaningful implications for the extent to which differences in choices across participant pools and time-cost environments are observable. Even when differences in choice quality exist, they should be attenuated in simple problems, where accuracy is already high, and in very complex problems, where choices are already noisy. At intermediate complexity, by contrast, further deliberation remains most productive. Differences in time costs should therefore generate the largest differences in deliberation and accuracy for such tasks.

Hypothesis 3. Differences in choice accuracy across participant pools and time-cost environments are largest at intermediate levels of task complexity.

Task complexity therefore provides a common test of the proposed mechanism across the baseline participant-pool comparison and both experimental interventions.

3. Experimental Design

We designed a four-treatment experiment to test Hypotheses 1, 2, and 3. The design combines two participant pools with two within-pool manipulations of the time-cost environment. The baseline treatments reproduce the conventional comparison between a university laboratory sample and a broader online sample, while the remaining treatments exogenously vary time costs within each participant pool.

3.1. Participants

We recruited 159 laboratory participants from the UCL Experimental Laboratory for Finance and Economics and 162 online participants from Prolific. In both cases, eligibility was restricted to adult participants located in the United Kingdom, fluent in English, and, for Prolific participants, with an approval rate of at least 95%. Both tasks included comprehension checks that participants had to pass before beginning the paid rounds, as well as a text captcha. Participants could only take part in the experiment once.

3.2. Treatments

The first two treatments implement the baseline environments represented by Lab Baseline and Online Baseline. In the **Lab Baseline** treatment, university students completed the experiment in the laboratory. In the **Online Baseline** treatment, participants recruited through Prolific completed the same experiment online. Consistent with the conceptual framework, laboratory participants remained in the laboratory until the end of the session, whereas online participants could leave the study immediately after completing it. The maximum duration was 75 minutes in the laboratory and 74 minutes online, the closest feasible match.[5] In both environments, payments were made within 24 hours and only after data collection had ended.

The remaining two treatments implement the time-cost manipulations represented by Lab Discounting and Online Delay. In **Lab Discounting**, the payoff from a correct answer was discounted as a function of response time at a rate of 3% per second. In **Online Delay**, participants faced a minimum response time of 30 seconds on each decision. They could select an answer before then but could not submit it and proceed to the next round until the delay expired. The treatment therefore did not reward accuracy directly; it reduced the opportunity cost of deliberating for longer during the first 30 seconds.

Treatment assignment was random within each participant pool. Laboratory participants were assigned with equal probability and evenly split to Lab Baseline or Lab Discounting in each session. Online participants were assigned with equal probability to Online Baseline or Online Delay and also evenly split. In addition to the 162 online participants that concluded the experiment, four

[5] Prolific requires researchers to specify a target completion time and determines the maximum permitted completion time from that target. A target duration of 40 minutes generated a maximum permitted duration of 74 minutes. No available target duration generated a maximum of exactly 75 minutes.

online participants did not proceed from the welcome page, two from each treatment; all the 159 laboratory participants concluded the experiment.

3.3. Tasks and Task Complexity

Participants completed two decision tasks with an objectively payoff-maximising alternative: a dots task and a sums task. Both tasks were adapted from Gonçalves et al. (2026).[6] Task order was randomised across participants. In the dots task, participants observed a box containing red and blue dots and earned £4.00 if they selected the colour that appeared more frequently. One colour always had exactly two more dots than the other. In the sums task, participants observed two arithmetic expressions and earned £4.00 if they selected the expression with the larger value. Each term was an integer drawn randomly from -10 to 10, subject to the constraint that the two expressions had different values. Figure 1 shows screenshots of the two tasks. In both tasks, each alternative was equally likely to be correct, and ties were excluded by construction.

We varied complexity within participant as in Gonçalves et al. (2026). The dots task used 4, 32, 64, or 256 total dots; the sums task used 2, 8, 16, or 64 terms per expression. We classify these levels as *very easy*, *medium easy*, *medium hard*, and *very hard*. The very easy problems were designed so that the correct answer could be identified with little deliberation. In the very hard problems, additional deliberation was expected to have relatively low value because identifying the correct answer within a reasonable period was difficult. The manipulation varied the return to deliberation while preserving an objective definition of choice accuracy.

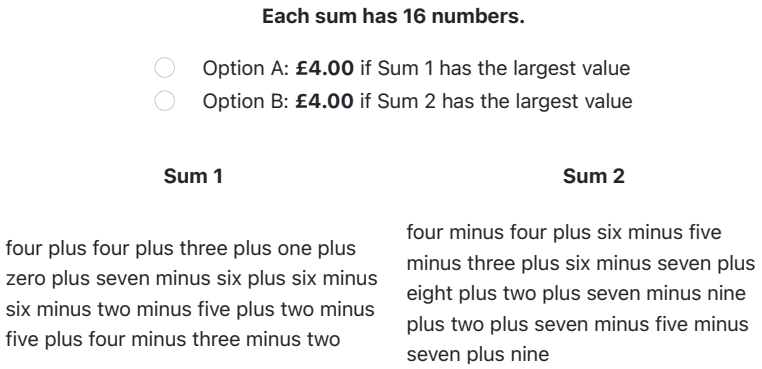
Each task began with instructions and comprehension questions that participants had to answer correctly before proceeding. First-attempt pass rates (75%) and average mistakes (0.32) did not differ statistically across participant pools. Participants then completed one unpaid practice round with feedback and 32 incentivised rounds without feedback. Task complexity varied randomly across rounds. Screenshots of the instructions and of the practice and paid rounds are reproduced in Online Appendix B.

3.4. Incentivisation and Payments

The incentive scheme, up to discounting in the Lab Discounting treatment, was the same across participant pools. One round from each task was selected uniformly at random for payment, consistent with standard practice in experiments involving multiple incentivised decisions (Charness et al., 2016; Azrieli et al., 2018). All participants received a fixed payment of £8.00 for completing the experiment. They could additionally earn either £0.00 or £4.00 from the randomly selected round in each task, depending on whether they chose the correct answer. Maximum total earnings before discounting were therefore £16.00. In Lab Discounting, the £4.00 bonus associated with a correct answer was exponentially discounted at a rate of 3% per second. The current value of the bonus was displayed on the decision screen and updated continuously as time elapsed. Average total earnings

[6]The experiment was coded using oTree (Chen et al., 2016).

(a) Dots



(b) Sums

Figure 1: Experimental Interface

Notes: The figure shows representative decision screens from the two tasks. In the dots task, participants receive a bonus for choosing the colour with most dots; in the sums task, for choosing the sum with the largest value.

were £13.97: £13.74 among laboratory participants and £14.20 among online participants.

4. Results

We present three sets of results: baseline participant-pool differences, effects of changing time costs within pool, and how both vary with task complexity.

For each comparison, we estimate variants of

$$y_{ij} = \beta_0 + \beta_1 D_i + \beta_2 \text{Sums}_{ij} + \beta_3 D_i \times \text{Sums}_{ij} + \varepsilon_{ij}, \tag{1}$$

where y_{ij} is either an indicator for participant i choosing the payoff-maximising alternative in decision j or the logarithm of response time in seconds. The indicator D_i identifies the participant

	Chose Correct		Log Time	
	(1)	(2)	(3)	(4)
Online Baseline	-0.085***	-0.077***	-0.427***	-0.305***
	(0.011)	(0.014)	(0.070)	(0.074)
Online Baseline x Sums	–	-0.016	–	-0.242**
		(0.020)		(0.108)
Sums	–	-0.011	–	0.773***
		(0.014)		(0.083)
Intercept	0.855***	0.861***	2.724***	2.337***
	(0.007)	(0.009)	(0.042)	(0.055)
R^2	0.01	0.01	0.03	0.12
N	10240	10240	10240	10240

Table 1: Choice Accuracy and Response Time by Participant Pool

Notes: The table reports linear regressions comparing Online Baseline with Lab Baseline. The dependent variable is an indicator for choosing the payoff-maximising alternative in columns (1)–(2) and log response time in columns (3)–(4). The task-specific specifications include an indicator for the sums task and its interaction with the online indicator; the omitted task is dots. Standard errors clustered at the participant level are reported in parentheses. *, **, and *** denote $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.

pool or time-cost treatment relevant to each comparison, and Sums_{ij} identifies decisions in the sums task. Specifications that pool the treatment effect across tasks omit the interaction term. We report decision-level linear regressions with standard errors clustered at the participant level. For log response time, we express treatment effects as $100\left[\exp(\widehat{\beta}) - 1\right]$ percent changes.

4.1. Baseline Differences Across Participant Pools

We begin with the baseline comparison between laboratory and online participants. Hypothesis 1 predicts that online participants choose less accurately and respond more quickly because they face a higher opportunity cost of time. Table 1 reports the corresponding differences, both pooled across tasks and allowing them to vary by task domain.

Online participants are 8.5 percentage points (p.p.) less likely to choose the payoff-maximising alternative than laboratory participants. Choice accuracy is 77.0% online and 85.5% in the laboratory. This difference appears in both task domains: accuracy is 78.4% online and 86.1% in the laboratory in the dots task, and 75.7% and 85.0%, respectively, in the sums task. The result reproduces the data-quality gap documented in previous participant-pool comparisons (Snowberg and Yariv, 2021; Gupta et al., 2021).

Online participants also respond substantially faster: response times are approximately 35% lower, with median response times of 8.4 seconds per round online and 13.9 seconds in the laboratory. The difference is again present in both tasks: median response times are 6.9 and 10.4 seconds in the dots task and 10.8 and 27.3 seconds in the sums task for online and laboratory participants, respectively.

The data also show that mean differences are not driven by a small number of participants. Participant-

level distributions are shifted toward lower accuracy and shorter response times online (Figure 3 and Table 6 in Online Appendix A).

Taken together, lower accuracy and shorter response times have the qualitative signature of higher time costs. The accuracy difference is consistent with existing evidence that online samples produce noisier choices. The response-time difference identifies the additional behavioural margin predicted by our framework. The baseline comparison alone, however, cannot separate the time-cost mechanism from fixed compositional differences between the two pools. The within-pool interventions provide a more direct test.

4.2. Discounting and Delay

We next test Hypothesis 2 by changing the time-cost environment while holding the participant pool fixed. Table 2 reports the effect of Lab Discounting relative to Lab Baseline and the effect of Online Delay relative to Online Baseline.

Table 2a shows that increasing the cost of time in the laboratory reduces both accuracy and response time. Discounting lowers average accuracy from 85.5% to 72.5%, a decline of 13.0 p.p. The log-response-time estimate implies a reduction in response time of approximately 68%. In levels, median response time falls from 13.9 seconds in Lab Baseline to 4.0 seconds under Lab Discounting. The effects have the same sign in both task domains.

The magnitude of the intervention more than eliminates the baseline laboratory advantage. Participants in Lab Discounting are less accurate and respond faster than participants in Online Baseline. This does not imply that the treatment exactly reproduces the online environment; rather, it shows that sufficiently strong time incentives can reverse the baseline ordering between the two participant pools.

Table 2b shows that lowering the opportunity cost of deliberation time improves online choice accuracy. Online Delay raises average accuracy from 77.0% to 81.8%, an increase of 4.9 p.p. The log-response-time estimate implies an increase in measured response time of approximately 275%. The corresponding median response time rises from 8.4 to 32.6 seconds.

Because the response-time increase is partly mechanical, the accuracy gain is the more informative behavioural response. Although the delay does not reward correct answers, preventing immediate progression improves choice quality. The effect is present in both task domains, although it is larger in the dots task, where baseline decisions are shorter and the 30-second constraint consequently binds more strongly.

The results are not attributable to a small subset of unusually responsive participants. The Online Appendix A shows that the average treatment effects correspond to broad shifts in participant-level accuracy distributions.

The two interventions move behaviour in opposite directions, as predicted by the framework. In-

	Lab			
	Chose Correct		Log Time	
	(1)	(2)	(3)	(4)
Lab Discounting	-0.130***	-0.126***	-1.124***	-0.939***
	(0.012)	(0.015)	(0.073)	(0.078)
Lab Discounting x Sums	–	-0.008	–	-0.370***
		(0.020)		(0.092)
Sums	–	-0.011	–	0.773***
		(0.014)		(0.083)
Intercept	0.855***	0.861***	2.724***	2.337***
	(0.007)	(0.009)	(0.042)	(0.055)
R^2	0.03	0.03	0.22	0.29
N	10176	10176	10176	10176

(a) Laboratory

	Online			
	Chose Correct		Log Time	
	(1)	(2)	(3)	(4)
Online Delay	0.048***	0.061***	1.323***	1.552***
	(0.012)	(0.014)	(0.058)	(0.053)
Online Delay x Sums	–	-0.027	–	-0.459***
		(0.020)		(0.074)
Sums	–	-0.027*	–	0.531***
		(0.014)		(0.069)
Intercept	0.770***	0.784***	2.298***	2.032***
	(0.009)	(0.011)	(0.056)	(0.049)
R^2	0.00	0.01	0.43	0.47
N	10368	10368	10368	10368

(b) Online

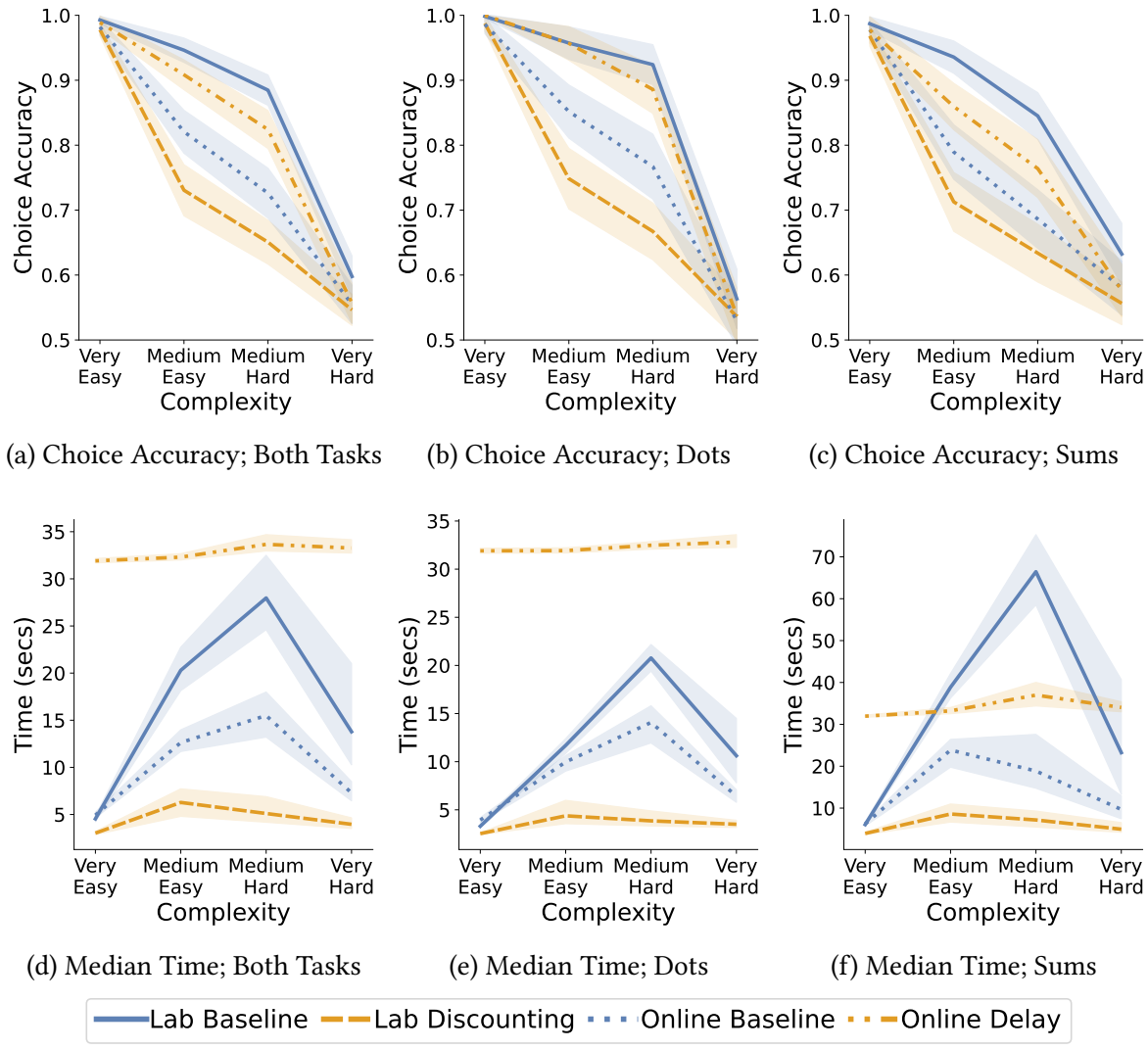
Table 2: Choice Accuracy and Response Time by Time-Cost Treatment

Notes: The table reports linear regressions comparing Lab Discounting and Online Delay with, respectively, Lab Baseline (panel (a)) and Online Baseline (panel (b)). The dependent variable is an indicator for choosing the payoff-maximising alternative in columns (1)–(2) and log response time in columns (3)–(4). The task-specific specifications include an indicator for the sums task and its interaction with the treatment indicator; the omitted task is dots. Standard errors clustered at the participant level are reported in parentheses. *, **, and *** denote $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.

creasing the opportunity cost of time in the laboratory produces faster and less accurate choices, whereas reducing it online increases accuracy. These results complement time-pressure experiments showing that decision time can affect measured behaviour (Kocher and Sutter, 2006; Baillon et al., 2018). Because the present comparisons hold the participant pool fixed and use objectively correct answers, they show directly that the timing environment affects decision quality rather than only measured preferences.

4.3. Task Complexity

Hypothesis 3 predicts that participant-pool and treatment differences should be largest at intermediate levels of complexity, where additional deliberation is most productive. Figure 2 presents choice accuracy and log response time by treatment, task domain, and complexity.



(a) Choice Accuracy; Both Tasks (b) Choice Accuracy; Dots (c) Choice Accuracy; Sums

(d) Median Time; Both Tasks (e) Median Time; Dots (f) Median Time; Sums

Figure 2: Choice Accuracy and Response Time by Time-Cost Treatment and Task Complexity

Notes: The figure plots choice accuracy and median response time by participant pool, time-cost treatment, task domain, and complexity level. Shaded areas correspond to 95% confidence intervals using standard errors clustered at the participant level.

The top row of Figure 2 shows a common pattern across the three comparisons. Differences in choice accuracy are negligible for very easy problems, largest at the two intermediate complexity levels, and substantially attenuated for very hard problems. For the baseline comparison, the laboratory advantage is 1.1 p.p. in very easy problems and not statistically significant. It rises to 12.6 p.p. at medium-easy complexity and 15.8 p.p. at medium-hard complexity (p-values< 0.001), before falling to 4.3 p.p. for very hard problems, again not statistically significant. Thus, the overall participant-pool difference documented in Table 1 is concentrated in the problems for which addi-

tional deliberation should have the greatest value. The effects of Lab Discounting and Online Delay follow the same pattern, with the largest accuracy changes at intermediate complexity levels.[7] The time-cost interventions therefore had little effect when the answer is nearly immediate or when further deliberation is ineffectual, but a large effect between these extremes.

The response-time patterns in the bottom row are also consistent with the framework. Outside Online Delay, where we directly impose a minimum response time, response times generally exhibit the inverse-U-shaped relationship with complexity documented by Gonçalves et al. (2026). Participants respond quickly to very easy problems, spend more time on intermediate problems, and reduce response time again for very hard problems.[8]

Online Delay raises measured response time at every complexity level because of the 30-second minimum. Its effect on accuracy, however, is concentrated at the two intermediate levels. This distinction is informative. Additional available time does not improve choices uniformly; it improves them when the task permits productive deliberation. The same complexity pattern appears in participant-level accuracy distributions (Figure 4 and Table 7 in Online Appendix A).

Taken together, the baseline pool gap, the effect of increasing laboratory time costs, and the effect of mitigating the opportunity cost of time by imposing a delay all vary with complexity in the manner predicted by the framework. The common pattern is difficult to explain using fixed differences in participant composition alone. It indicates that participant-pool differences depend jointly on the time-cost environment and on the value of deliberation in the task.

5. Discussion

Participant-pool comparisons are usually interpreted as comparisons between different kinds of people. Our results show that they are also comparisons between decision environments. At baseline, online participants respond more quickly and choose less accurately than laboratory participants. Holding the participant pool fixed, increasing the opportunity cost of time in the laboratory reduces response time and accuracy, while reducing it online improves accuracy. These effects, like the baseline gap, are largest at intermediate complexity, where additional deliberation is most productive. Together, the findings identify the opportunity cost of deliberation as a mechanism capable of generating, amplifying, and mitigating participant-pool differences.

This does not imply that time costs explain all laboratory–online differences. The baseline samples differ in composition, and our interventions do not reproduce the exact counterfactual environment faced by the other pool. Indeed, differences in time costs may be intrinsically tied to the fact that participants recruited from online platforms can choose from a multitude of alternative paid studies.

[7] The effects are concentrated at intermediate complexity: discounting lowers accuracy between 21.6–23.4 p.p. at the two intermediate levels (p-values< 0.001), compared with 1.5 and 5.1 p.p. at the extremes. Delay raises accuracy by 8.8 and 9.8 p.p. at the intermediate levels (p-values< 0.001), compared with 0.7 and 0.2 p.p. at the extremes.

[8] These patterns do not seem driven by participants simply quitting since the participant-level distribution of response times are faster still for very easy problems compared to very hard ones.

What the experiment establishes is that behaviour commonly attributed to participant composition changes when the timing environment changes while composition is held fixed. Fixed differences in preferences, ability, experience, or decision rules cannot alone account for the within-pool treatment effects. Observed behaviour is therefore jointly determined by participant characteristics and the elicitation environment.

This distinction matters for external validity. Changing participant pools is generally a compound treatment: it changes both who participates and the conditions under which they decide (Levitt and List, 2007; Harrison and List, 2004; Charness et al., 2013). A laboratory participant who cannot leave after an individual decision faces a different problem from an online participant who can immediately begin another activity. Even observing the same person online and in the laboratory may not reproduce the environment faced by participants recruited through platforms with readily available paid studies. The relevant question is therefore not whether one participant pool is generically superior (Snowberg and Yariv, 2021; Gupta et al., 2021), but whether the combination of participants, incentives, constraints, and task complexity approximates the target economic environment.

The complexity results highlight that participant-pool comparisons are local to the tasks on which they are measured. Very easy problems can produce similar behaviour because accuracy is near its ceiling, while very difficult problems can do so because further deliberation has little value. Neither null implies that participant composition or the decision environment would be irrelevant where information processing has a higher return, and a sizeable gap in one task need not generalise to tasks of different complexity. This may help reconcile heterogeneous findings across participant pools (Paolacci et al., 2010; Horton et al., 2011; Snowberg and Yariv, 2021; Gupta et al., 2021). What matters is not simply whether one task is more complex than another, but whether additional deliberation remains productive at the relevant complexity (Oprea, 2025; Gonçalves, 2026).

The same logic applies to measures of data quality. Comprehension checks, dominated alternatives, and other objectively scored choices are informative only to the extent that their complexity resembles that of the substantive decisions (Berinsky et al., 2014; Çelebi et al., 2026). A simple check may conceal noise in a more demanding outcome, while an excessively difficult check may make a sample appear unreliable for comparatively simple choices. This concern is especially important when the main outcomes have no objectively correct or optimal choice. Choice noise can be mistaken for preference heterogeneity and generate spurious relationships between elicited preferences and participant characteristics (Andersson et al., 2016). Time costs may likewise affect the precision with which preferences are expressed or the procedures used to form choices (Baillon et al., 2018; Imas et al., 2022), and differences in elicited preferences or beliefs may therefore combine genuine heterogeneity with environment-induced decision noise.

The Online Delay treatment provides a proof of concept that data quality is responsive to experimental design. A minimum response time can improve accuracy without changing the reward for a correct answer by promoting additional deliberation. Such delays may be useful in online ex-

periments and surveys, where respondents may otherwise minimise cognitive effort or optimise completion speed (Krosnick, 1991; Malhotra, 2008). Their effectiveness is not mechanical: the delay raises measured response time at every complexity level but improves accuracy mainly where further deliberation is productive (Andersen et al., 2018).

This implication extends beyond economic experiments. Response times can reveal preference intensity, confidence, or resolve (Tourangeau et al., 2000; Clithero, 2018; Alós-Ferrer et al., 2021; Liu and Netzer, 2023), but they need not be treated only as passive outcomes. A short page-progression delay may improve data quality in economic-sentiment surveys, political polls, and related settings by reducing the opportunity cost of forming a considered response. Delays can also be diagnostic: if a delay raises objective accuracy, rapid completion was contributing to measurement error; if it raises response time without improving accuracy, further deliberation may have little value (Krosnick, 1991; Yan and Tourangeau, 2008; Malhotra, 2008).

Our evidence is specific to two tasks with objective payoff-maximising choices, two participant pools, and particular timing interventions. The results do not imply that delays improve every outcome or that online participants generally deliberate too little. They establish that decision quality is endogenous to the timing environment and that this relationship further depends on task complexity. Future work can examine whether the mechanism extends to preference elicitation, belief formation, strategic reasoning, and the use of heuristics.

6. References

Agranov, M., A. Schotter, and I. Trevino. 2025. "Complex for Whom? An Experimental Approach to Subjective Complexity." *Working Paper* 1–51.

Alós-Ferrer, C., E. Fehr, and N. Netzer. 2021. "Time Will Tell: Recovering Preferences When Choices Are Noisy." *Journal of Political Economy* 129 (6): 1828–1877. 10.1086/713732.

Andersen, S., U. Gneezy, A. Kajackaite, and J. Marx. 2018. "Allowing for reflection time does not change behavior in dictator and cheating games." *Journal of Economic Behavior & Organization* 145 24–33. 10.1016/j.jebo.2017.10.012.

Andersson, O., H. J. Holm, J.-R. Tyran, and E. Wengström. 2016. "Risk Aversion Relates to Cognitive Ability: Preferences Or Noise?" *Journal of the European Economic Association* 14 (5): 1129–1154. 10.1111/jeea.12179.

Azrieli, Y., C. P. Chambers, and P. J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126 (4): 1472–1503. 10.1086/698136.

Baillon, A., Z. Huang, A. Selim, and P. P. Wakker. 2018. "Measuring Ambiguity Attitudes for All (Natural) Events." *Econometrica* 86 (5): 1839–1858. 10.3982/ECTA14370.

Berinsky, A. J., M. F. Margolis, and M. W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58 (3): 739–753. 10.1111/ajps.12081.

Bossaerts, P., and C. Murawski. 2017. "Computational Complexity and Human Decision-Making." *Trends in Cognitive Sciences* 21 (12): 917–929. 10.1016/j.tics.2017.09.005.

Çelebi, C., C. Exley, S. Harrs, H. Kivimaki, M. Serra-Garcia, and J. Yusof. 2026. "Mission Possible: The Collection of High-Quality Data." *CESifo Working Paper* 12535 1–22. 10.65864/q5ekls6q5u.

Charness, G., U. Gneezy, and B. Halladay. 2016. "Experimental methods: Pay one or pay all." *Journal of Economic Behavior & Organization* 131 (A): 141–150. 10.1016/j.jebo.2016.08.010.

Charness, G., U. Gneezy, and M. A. Kuhn. 2013. "Experimental Methods: Extra-Laboratory Experiments—Extending the Reach of Experimental Economics." *Journal of Economic Behavior & Organization* 91 93–100. 10.1016/j.jebo.2013.04.002.

Chen, D. L., M. Schonger, and C. Wickens. 2016. "oTree — An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance* 9 88–97. 10.1016/j.jbef.2015.12.001.

Chiong, K., M. Shum, R. Webb, and R. Chen. 2023. "Combining Choice and Response Time Data: A Drift-Diffusion Model of Mobile Advertisements." *Management Science* 1–20. 10.1287/mnsc.2023.4738.

Clithero, J. 2018. "Improving out-of-sample predictions using response times and a model of the decision process." *Journal of Economic Behavior and Organization* 148 344–375. 10.1016/j.jebo.2018.02.007.

Enke, B., and T. Graeber. 2023. "Cognitive Uncertainty." *Quarterly Journal of Economics* 138 (4): 2021–2067. 10.1093/qje/qjad025.

Enke, B., T. Graeber, R. Oprea, and J. Yang. 2026. "Behavioral Attenuation." *Working Paper* 1–162.

Fehr, E., and A. Rangel. 2011. "Neuroeconomic Foundations of Economic Choice—Recent Advances." *Journal of Economic Perspectives* 25 (4): 3–30. 10.1257/jep.25.4.3.

Forstmann, B., R. Ratcliff, and E.-J. Wagenmakers. 2016. "Sequential Sampling Models in Cog-

nitive Neuroscience: Advantages, Applications, and Extensions." *Annual Review of Psychology* 67 (1): 641–666. 10.1146/annurev-psych-122414-033645.

Fromell, H., D. Nosenzo, and T. Owens. 2020. "Altruism, fast and slow? Evidence from a meta-analysis and a new experiment." *Experimental Economics* 23 (4): 979–1001. 10.1007/s10683-020-09645-z.

Fréchette, G. R. 2017. "7. Experimental Economics across Subject Populations." In *The Handbook of Experimental Economics, Volume 2*, edited by Kagel, J. H., and A. E. Roth 141–165, Princeton University Press, . 10.1515/9781400883172-008.

Fudenberg, D., P. Strack, and T. Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review* 108 (12): 3651–3684. 10.1257/aer.20150742.

Gonçalves, D., S. Nunnari, and J. Zarate-Pina. 2026. "Revealing Complexity." *Working Paper*.

Gonçalves, D. 2026. "Speed, Accuracy, and Complexity." *Working Paper*.

Gupta, N., L. Rigotti, and A. Wilson. 2021. "The Experimenters' Dilemma: Inferential Preferences over Populations." *arXiv*. 10.48550/arXiv.2107.05064.

Haji, A. E., M. Krawczyk, M. Sylwestrzak, and E. Zawojska. 2019. "Time pressure and risk taking in auctions: A field experiment." *Journal of Behavioral and Experimental Economics* 78 68–79. 10.1016/j.socec.2018.12.001.

Harrison, G. W., and J. A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–1055. 10.1257/0022051043004577.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser. 2011. "The online laboratory: Conducting experiments in a real labor market." *Experimental Economics* 14 (3): 399–425. 10.1007/s10683-011-9273-9.

Imas, A., M. A. Kuhn, and V. Mironova. 2022. "Waiting to Choose: The Role of Deliberation in Intertemporal Choice." *American Economic Journal: Microeconomics* 14 (3): 414–440. 10.1257/mic.20180233.

Kirchler, M., D. Andersson, C. Bonn, M. Johannesson, E. Ø. Sørensen, M. Stefan, G. Tinghög, and D. Västfjäll. 2017. "The effect of fast and slow decisions on risk taking." *Journal of Risk and Uncertainty* 54 (1): 37–59. 10.1007/s11166-017-9252-4.

Kocher, M. G., J. Pahlke, and S. T. Trautmann. 2013. "Tempus fugit: Time pressure in risky decisions." *Management Science* 59 (10): 2380–2391. 10.1287/mnsc.2013.1711.

Kocher, M. G., and M. Sutter. 2006. "Time is money—Time pressure, incentives, and the quality of decision-making." *Journal of Economic Behavior & Organization* 61 (3): 375–392. 10.1016/j.jebo.2004.11.013.

Krajbich, I., K. Fernandez, and X. Yang. 2026. "Decomposing Economic Choices with Drift-Diffusion Models." In *Neuroeconomics: Core Topics and Current Directions*, edited by Smith, D. V., P. L. Lockwood, and D. S. Fareri 141–165, Springer, . 10.1007/978-3-032-02925-6_9.

Krosnick, J. A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5 (3): 213–236. 10.1002/acp.2350050305.

Levitt, S. D., and J. A. List. 2007. "On the generalizability of lab behaviour to the field." *Canadian Journal of Economics/Revue canadienne d'économique* 40 (2): 347–370. 10.1111/j.1365-2966.2007.00412.x.

Liu, S., and N. Netzer. 2023. "Happy Times: Measuring Happiness Using Response Times." *American Economic Review* 113 (12): 3289–3322. 10.1257/aer.20211051.

Malhotra, N. 2008. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion*

Quarterly 72 (5): 914–934. 10.1093/poq/nfn050.

Oprea, R. 2020. "What Makes a Rule Complex?" *American Economic Review* 110 (12): 3913–3951. 10.1257/aer.20191717.

Oprea, R. 2025. "Complexity and its Measurement." *Working Paper* 1–65.

Paolacci, G., J. Chandler, and P. G. Ipeirotis. 2010. "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5 (5): 411–419. 10.1017/S1930297500002205.

Ray, D. ⓡ A. Robson. 2018. "Certified Random: A New Order for Coauthorship." *American Economic Review* 108 (2): 489–520. 10.1257/aer.20161492.

Sanjurjo, A. 2025. "Space Complexity in Choice." *Working Paper* 1–156.

Schotter, A., and I. Trevino. 2021. "Is response time predictive of choice? An experimental study of threshold strategies." *Experimental Economics* 24 87–117. 10.1007/s10683-020-09651-1.

Snowberg, E., and L. Yariv. 2021. "Testing the waters: Behavior across participant pools." *American Economic Review* 111 (2): 687–719. 10.1257/aer.20181065.

Spiliopoulos, L., and A. Ortmann. 2018. "The BCD of response time analysis in experimental economics." *Experimental Economics* 21 383–433. 10.1007/s10683-017-9528-1.

Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Yan, T., and R. Tourangeau. 2008. "Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology* 22 (1): 51–68. 10.1002/acp.1331.

Online Appendix A. Additional Figures and Tables

	Log Time	
	(1)	(2)
Online Baseline	-0.090	-0.056
	(0.063)	(0.076)
Online Baseline x Sums	–	-0.070
		(0.119)
Sums	–	-0.959*
		(0.555)
Participant Avg Accuracy	3.938***	2.926***
	(0.436)	(0.478)
Participant Avg Accuracy x Sums	–	2.025***
		(0.626)
Intercept	-0.645*	-0.165
	(0.376)	(0.415)
R^2	0.10	0.18
N	10240	10240

Table 3: Response Time by Participant Pool, Conditional on Choice Accuracy

Notes: The table reports linear regressions of log response time on an indicator for Online Baseline and participant-level average choice accuracy. Column (2) additionally includes an indicator for the sums task and its interactions with the online indicator and participant-level average accuracy. The omitted category is Lab Baseline in the dots task. Standard errors clustered at the participant level are reported in parentheses. *, **, and *** denote $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.

	Chose Correct	Log Time
	(1)	(2)
Online Baseline	-0.011	0.106**
	(0.007)	(0.043)
Online Baseline x Medium Easy	-0.115***	-0.553***
	(0.018)	(0.058)
Online Baseline x Medium Hard	-0.148***	-0.830***
	(0.023)	(0.094)
Online Baseline x Very Hard	-0.032	-0.745***
	(0.023)	(0.130)
Medium Easy	-0.046***	1.465***
	(0.009)	(0.032)
Medium Hard	-0.108***	1.802***
	(0.012)	(0.060)
Very Hard	-0.395***	1.359***
	(0.016)	(0.101)
Intercept	0.993***	1.564***
	(0.003)	(0.026)
R^2	0.17	0.26
N	10240	10240

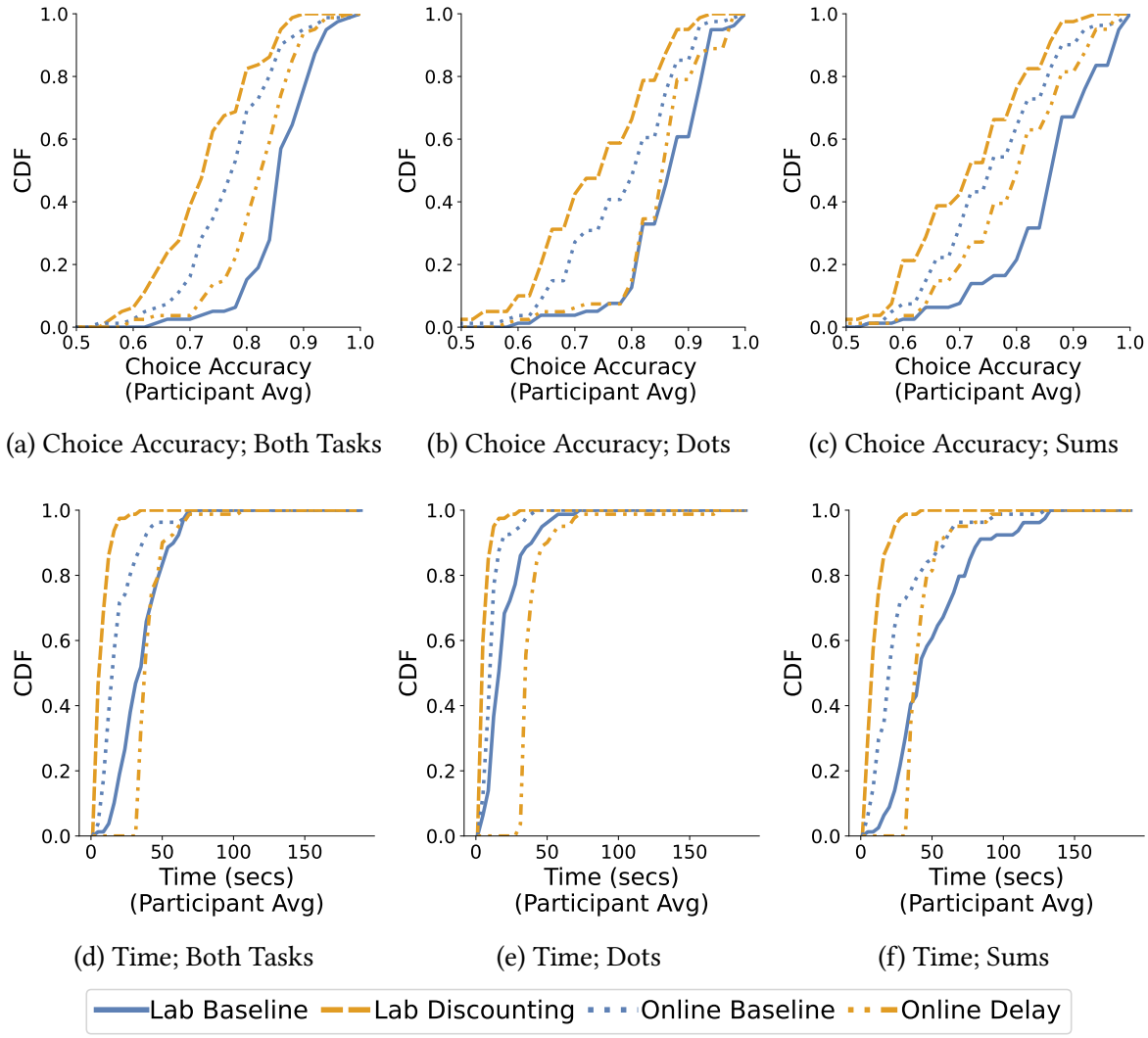
Table 4: Baseline Choice Accuracy and Response Time by Task Complexity

Notes: The table reports linear regressions comparing Online Baseline with Lab Baseline across task-complexity levels. The dependent variable is an indicator for choosing the payoff-maximising alternative in column (1) and log response time in column (2). The specifications include indicators for task complexity and their interactions with the online indicator. The omitted category is Lab Baseline in very easy problems. Standard errors clustered at the participant level are reported in parentheses. *, **, and *** denote $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.

	Pooled		Lab		Online	
	Chose Correct	Log Time	Chose Correct	Log Time	Chose Correct	Log Time
	(1)	(2)	(3)	(4)	(5)	(6)
Online Baseline	-0.011	0.106**	–	–	–	–
	(0.007)	(0.043)				
Lab Discounting	-0.015***	-0.389***	-0.015***	-0.389***	–	–
	(0.006)	(0.035)	(0.006)	(0.035)		
Online Delay	-0.004	1.969***	–	–	0.007	1.863***
	(0.006)	(0.028)			(0.008)	(0.036)
Online Baseline x Medium Easy	-0.115***	-0.553***	–	–	–	–
	(0.018)	(0.058)				
Lab Discounting x Medium Easy	-0.201***	-0.809***	-0.201***	-0.809***	–	–
	(0.021)	(0.071)	(0.021)	(0.071)		
Online Delay x Medium Easy	-0.034**	-1.406***	–	–	0.081***	-0.853***
	(0.014)	(0.033)			(0.019)	(0.049)
Online Baseline x Medium Hard	-0.148***	-0.830***	–	–	–	–
	(0.023)	(0.094)				
Lab Discounting x Medium Hard	-0.219***	-1.187***	-0.219***	-1.187***	–	–
	(0.021)	(0.094)	(0.021)	(0.094)		
Online Delay x Medium Hard	-0.057***	-1.670***	–	–	0.091***	-0.840***
	(0.019)	(0.062)			(0.025)	(0.075)
Online Baseline x Very Hard	-0.032	-0.745***	–	–	–	–
	(0.023)	(0.130)				
Lab Discounting x Very Hard	-0.037*	-0.934***	-0.037*	-0.934***	–	–
	(0.021)	(0.123)	(0.021)	(0.123)		
Online Delay x Very Hard	-0.037	-1.204***	–	–	-0.005	-0.459***
	(0.023)	(0.104)			(0.023)	(0.087)
Medium Easy	-0.046***	1.465***	-0.046***	1.465***	-0.161***	0.912***
	(0.009)	(0.032)	(0.009)	(0.032)	(0.016)	(0.049)
Medium Hard	-0.108***	1.802***	-0.108***	1.802***	-0.256***	0.972***
	(0.012)	(0.060)	(0.012)	(0.060)	(0.020)	(0.073)
Very Hard	-0.395***	1.359***	-0.395***	1.359***	-0.427***	0.614***
	(0.016)	(0.100)	(0.016)	(0.101)	(0.016)	(0.083)
Intercept	0.993***	1.564***	0.993***	1.564***	0.982***	1.670***
	(0.003)	(0.026)	(0.003)	(0.026)	(0.006)	(0.034)
R^2	0.17	0.51	0.17	0.41	0.16	0.51
N	20544	20544	10176	10176	10368	10368

Table 5: Choice Accuracy and Response Time by Time-Cost Treatment and Task Complexity

Notes: The table reports decision-level linear regressions of choice accuracy and log response time on treatment indicators, task-complexity indicators, and their interactions. Columns (1)–(2) use all treatments, columns (3)–(4) restrict the sample to laboratory participants, and columns (5)–(6) restrict the sample to online participants. The omitted category is Lab Baseline in very easy problems in columns (1)–(4) and Online Baseline in very easy problems in columns (5)–(6). Standard errors clustered at the participant level are reported in parentheses. *, **, and *** denote $p < 0.1$, $p < 0.05$, and $p < 0.01$, respectively.

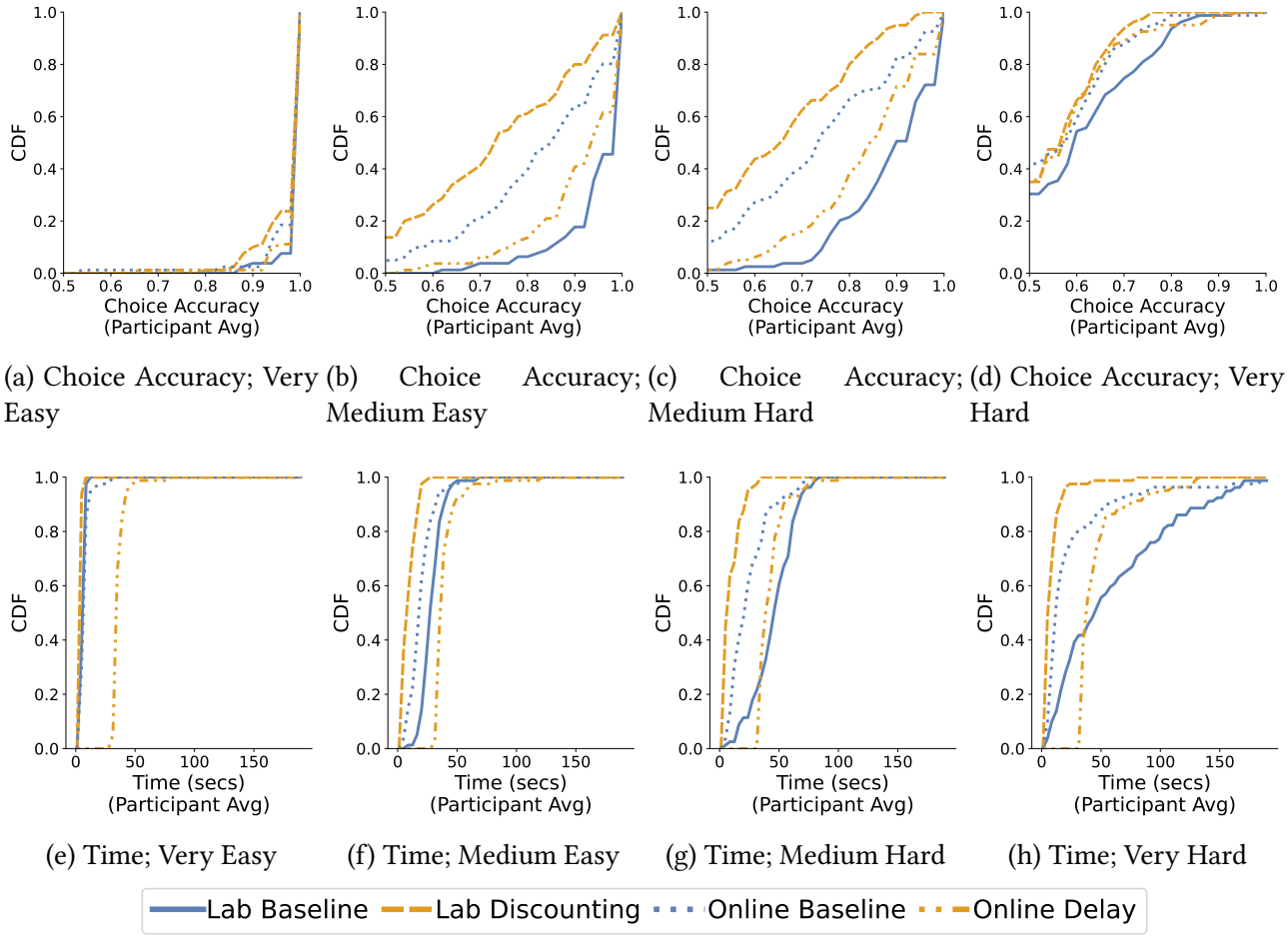


(a) Choice Accuracy; Both Tasks (b) Choice Accuracy; Dots (c) Choice Accuracy; Sums

(d) Time; Both Tasks (e) Time; Dots (f) Time; Sums

Figure 3: Participant-Level Choice Accuracy and Response Time Distributions by Treatment and Task

Notes: The figure plots empirical cumulative distribution functions of participant-level average choice accuracy in the top row and participant-level average response time, in seconds, in the bottom row. The left column pools both tasks; the middle and right columns restrict the data to the dots and sums tasks, respectively.



(a) Choice Accuracy; Very Easy (b) Choice Accuracy; Medium Easy (c) Choice Accuracy; Medium Hard (d) Choice Accuracy; Very Hard

(e) Time; Very Easy (f) Time; Medium Easy (g) Time; Medium Hard (h) Time; Very Hard

Figure 4: Participant-Level Choice Accuracy and Response Time Distributions by Treatment and Task Complexity

Notes: The figure plots empirical cumulative distribution functions of participant-level average choice accuracy in the top row and participant-level average response time, in seconds, in the bottom row. From left to right, the columns restrict the data to very easy, medium-easy, medium-hard, and very hard problems.

Table 6: First-Order Stochastic Dominance Tests by Treatment and Task

(a) Choice Accuracy; Both Tasks

		Row $\succeq_{FOSD}$ Column			
		Lab		Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.01	0.00
	Discounting	0.67^{***}	–	0.28^{***}	0.53^{***}
Online	Baseline	0.54^{***}	0.01	–	0.37^{***}
	Delay	0.31^{***}	0.00	0.01	–

(b) Time; Both Tasks

		Row $\succeq_{FOSD}$ Column			
		Lab		Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.01	0.48^{***}
	Discounting	0.86^{***}	–	0.54^{***}	0.99^{***}
Online	Baseline	0.55^{***}	0.01	–	0.85^{***}
	Delay	0.08	0.00	0.00	–

(c) Choice Accuracy; Dots

		Row $\succeq_{FOSD}$ Column			
		Lab		Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.06
	Discounting	0.54^{***}	–	0.18^{*}	0.51^{***}
Online	Baseline	0.35^{***}	0.00	–	0.33^{***}
	Delay	0.18^{*}	0.00	0.01	–

(d) Time; Dots

		Row $\succeq_{FOSD}$ Column			
		Lab		Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.86^{***}
	Discounting	0.74^{***}	–	0.53^{***}	1.00^{***}
Online	Baseline	0.42^{***}	0.00	–	0.95^{***}
	Delay	0.00	0.00	0.00	–

(e) Choice Accuracy; Sums

		Row $\succeq_{FOSD}$ Column			
		Lab		Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.00
	Discounting	0.55^{***}	–	0.17^{*}	0.27^{***}
Online	Baseline	0.43^{***}	0.00	–	0.16
	Delay	0.31^{***}	0.01	0.02	–

(f) Time; Sums

		Row $\succeq_{FOSD}$ Column			
		Lab		Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.30^{***}
	Discounting	0.82^{***}	–	0.53^{***}	0.99^{***}
Online	Baseline	0.53^{***}	0.01	–	0.74^{***}
	Delay	0.28^{***}	0.00	0.04	–

Notes: The table reports the one-sided Kolmogorov–Smirnov statistic for the null hypothesis that the participant-level distribution for the row treatment first-order stochastically dominates that for the column treatment. Choice accuracy is the participant-level frequency of payoff-maximising choices; response time is the participant-level average response time in seconds. Panels (a)–(b) pool both tasks, panels (c)–(d) restrict the data to the dots task, and panels (e)–(f) restrict the data to the sums task. *, **, and *** denote rejection of the dominance null with p-values < 0.1, < 0.05, and < 0.01, respectively.

Table 7: First-Order Stochastic Dominance Tests by Treatment and Task Complexity

(a) Choice Accuracy; Very Easy

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.01	0.03
	Discounting	0.16	–	0.09	0.13
Online	Baseline	0.11	0.01	–	0.07
	Delay	0.06	0.01	0.00	–

(b) Choice Accuracy; Medium Easy

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.00
	Discounting	0.65^{***}	–	0.28^{***}	0.49^{***}
Online	Baseline	0.49^{***}	0.00	–	0.32^{***}
	Delay	0.27^{***}	0.00	0.00	–

(c) Choice Accuracy; Medium Hard

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.00
	Discounting	0.61^{***}	–	0.24^{***}	0.49^{***}
Online	Baseline	0.45^{***}	0.04	–	0.33^{***}
	Delay	0.23^{**}	0.00	0.00	–

(d) Choice Accuracy; Very Hard

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.04	0.01	0.04
	Discounting	0.20^{**}	–	0.07	0.07
Online	Baseline	0.16	0.09	–	0.06
	Delay	0.18^{*}	0.04	0.06	–

(e) Time; Very Easy

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.13	1.00^{***}
	Discounting	0.66^{***}	–	0.78^{***}	1.00^{***}
Online	Baseline	0.01	0.01	–	1.00^{***}
	Delay	0.00	0.00	0.00	–

(f) Time; Medium Easy

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.01	0.65^{***}
	Discounting	0.87^{***}	–	0.55^{***}	1.00^{***}
Online	Baseline	0.50^{***}	0.00	–	0.88^{***}
	Delay	0.00	0.00	0.00	–

(g) Time; Medium Hard

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.00	0.22^{**}
	Discounting	0.84^{***}	–	0.50^{***}	0.98^{***}
Online	Baseline	0.54^{***}	0.01	–	0.72^{***}
	Delay	0.30^{***}	0.00	0.00	–

(h) Time; Very Hard

		Row $\geq_{FOSD}$ Column			
			Lab	Online	
		Baseline	Discounting	Baseline	Delay
	(1)	(2)	(3)	(4)	
Lab	Baseline	–	0.00	0.01	0.42^{***}
	Discounting	0.74^{***}	–	0.41^{***}	0.98^{***}
Online	Baseline	0.45^{***}	0.00	–	0.80^{***}
	Delay	0.28^{***}	0.00	0.04	–

Notes: The table reports the one-sided Kolmogorov–Smirnov statistic for the null hypothesis that the participant-level distribution for the row treatment first-order stochastically dominates that for the column treatment. Choice accuracy is the participant-level frequency of payoff-maximising choices; response time is the participant-level average response time in seconds. Panels (a) and (e), (b) and (f), (c) and (g), and (d) and (h) restrict the data to very easy, medium-easy, medium-hard, and very hard problems, respectively. *, **, and *** denote rejection of the dominance null with p-values < 0.1, < 0.05, and < 0.01, respectively.

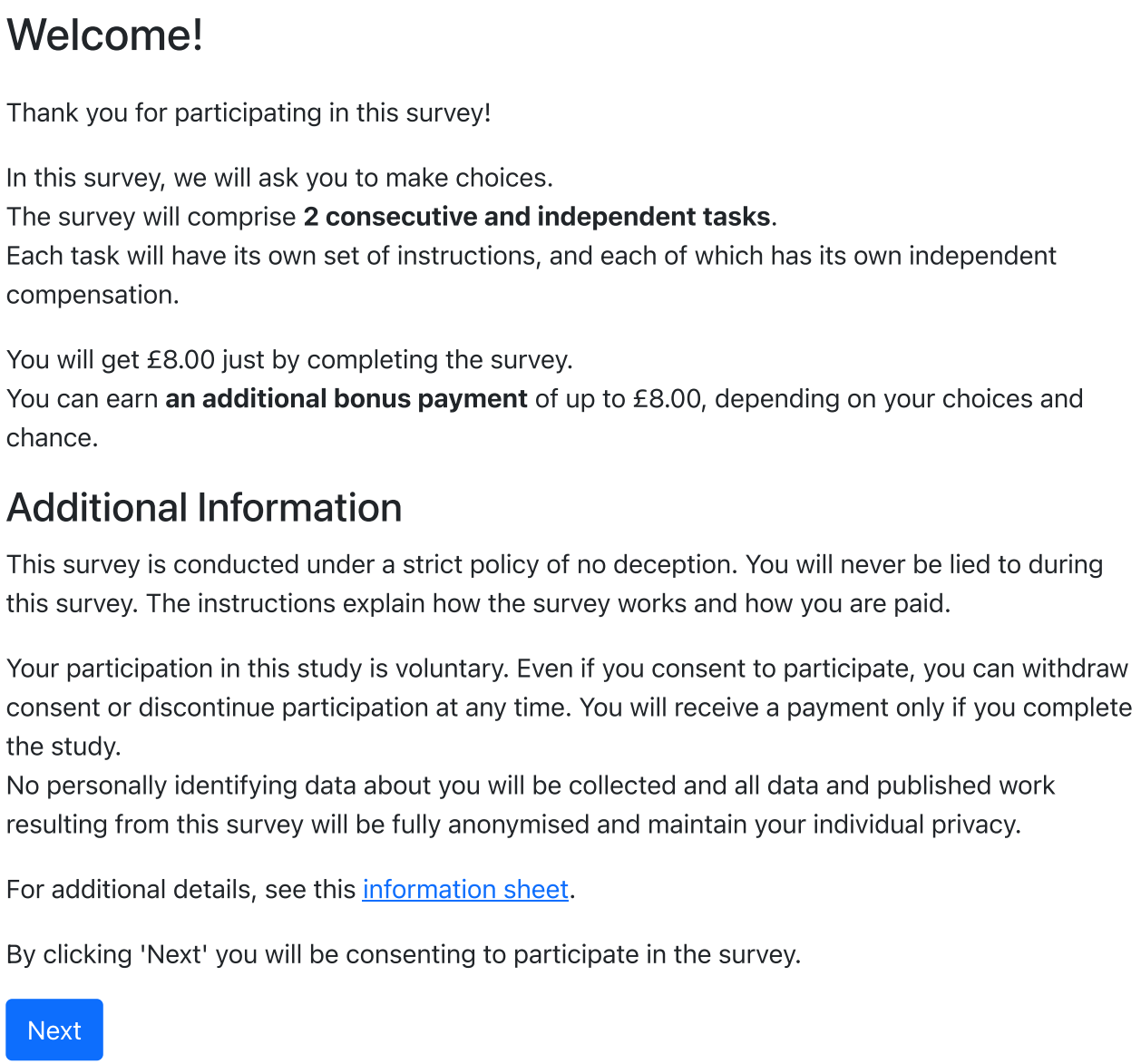
Online Appendix B. Instructions and Interface

Below we reproduce screenshots with the instructions, practice rounds, and incentivised rounds.

We start with the baseline treatments.

Online Appendix B.1. Welcome Screen

Lab Baseline.



Welcome!

Thank you for participating in this survey!

In this survey, we will ask you to make choices.
The survey will comprise **2 consecutive and independent tasks**.
Each task will have its own set of instructions, and each of which has its own independent compensation.

You will get £8.00 just by completing the survey.
You can earn **an additional bonus payment** of up to £8.00, depending on your choices and chance.

Additional Information

This survey is conducted under a strict policy of no deception. You will never be lied to during this survey. The instructions explain how the survey works and how you are paid.

Your participation in this study is voluntary. Even if you consent to participate, you can withdraw consent or discontinue participation at any time. You will receive a payment only if you complete the study.
No personally identifying data about you will be collected and all data and published work resulting from this survey will be fully anonymised and maintain your individual privacy.

For additional details, see this information sheet.

By clicking 'Next' you will be consenting to participate in the survey.

Next

Welcome!

Thank you for participating in this survey!

In this survey, we will ask you to make choices.
The survey will comprise **2 consecutive and independent tasks**.
Each task will have its own set of instructions, and each of which has its own independent compensation.

You will get £8.00 just by completing the survey.
You can earn **an additional bonus payment** of up to £8.00, depending on your choices and chance.

'Bot'-Detection

This survey is designed for humans and cannot be fulfilled using automated answers.

You will be asked to prove you are complying with this requirement by transcribing words at random points in this survey. The text will be as legible as the text in these instructions. Any human able to read this text will be able to read the words for transcription, but a 'bot' will not. You will be allowed 3 attempts and 1 minute per attempt. If you fail to transcribe a word three times, the survey will be immediately terminated and you will not be able to complete it nor receive payment.

Additional Information

This survey is conducted under a strict policy of no deception. You will never be lied to during this survey. The instructions explain how the survey works and how you are paid.

Your participation in this study is voluntary. Even if you consent to participate, you can withdraw consent or discontinue participation at any time. You will receive a payment only if you complete the study.
No personally identifying data about you will be collected and all data and published work resulting from this survey will be fully anonymised and maintain your individual privacy.

For additional details, see this information sheet.

By clicking 'Next' you will be consenting to participate in the survey.

Next

Welcome!

Thank you for participating in this survey!

In this survey, we will ask you to make choices.
The survey will comprise **2 consecutive and independent tasks**.
Each task will have its own set of instructions, and each of which has its own independent compensation.

You will get £8.00 just by completing the survey.
You can earn **an additional bonus payment** of up to £8.00, depending on your choices and chance.

Your bonus will be discounted by 3.0% for each second you take.
This means that, for example, if you take 10 seconds in a particular task, you would get about 26.25% less of whichever bonus you get from that task.
Time spent on a particular task will not affect the bonus from other tasks.

Additional Information

This survey is conducted under a strict policy of no deception. You will never be lied to during this survey. The instructions explain how the survey works and how you are paid.

Your participation in this study is voluntary. Even if you consent to participate, you can withdraw consent or discontinue participation at any time. You will receive a payment only if you complete the study.
No personally identifying data about you will be collected and all data and published work resulting from this survey will be fully anonymised and maintain your individual privacy.

For additional details, see this information sheet.

By clicking 'Next' you will be consenting to participate in the survey.

Next

Welcome!

Thank you for participating in this survey!

In this survey, we will ask you to make choices.
The survey will comprise **2 consecutive and independent tasks**.
Each task will have its own set of instructions, and each of which has its own independent compensation.

You will get £8.00 just by completing the survey.
You can earn **an additional bonus payment** of up to £8.00, depending on your choices and chance.

You will need to wait a few seconds before you can submit each answer.

'Bot'-Detection

This survey is designed for humans and cannot be fulfilled using automated answers.

You will be asked to prove you are complying with this requirement by transcribing words at random points in this survey. The text will be as legible as the text in these instructions. Any human able to read this text will be able to read the words for transcription, but a 'bot' will not. You will be allowed 3 attempts and 1 minute per attempt. If you fail to transcribe a word three times, the survey will be immediately terminated and you will not be able to complete it nor receive payment.

Additional Information

This survey is conducted under a strict policy of no deception. You will never be lied to during this survey. The instructions explain how the survey works and how you are paid.

Your participation in this study is voluntary. Even if you consent to participate, you can withdraw consent or discontinue participation at any time. You will receive a payment only if you complete the study.
No personally identifying data about you will be collected and all data and published work resulting from this survey will be fully anonymised and maintain your individual privacy.

For additional details, see this information sheet.

By clicking 'Next' you will be consenting to participate in the survey.

Next

Online Appendix B.2. Instructions

Lab Baseline and Online Baseline.

Sums Task

Task 2 - Instructions

This task will have 32 rounds.
At the end of the task, one round will be chosen randomly and what you earned in that round will be added to your total earnings.
All rounds are equally likely to be selected.

For each round of this survey, you will be shown two sums, Sum 1 and Sum 2.
Each sum comprises many randomly selected integers ranging from 0 to 9.
Each of the sums is equally likely to have the largest value; there are no ties.

In each round, you will be asked to choose between two options: Option A and Option B.
Each option pays £4.00 if it indicates the sum with the largest value.

For example, you could be asked to choose between the following options:

Option A: **£4.00** if Sum 1 has the largest value

Option B: **£4.00** if Sum 2 has the largest value

In order to get familiarised with the task, you will begin with a practice round. You will not be paid for the practice round.

Questions

You must answer the following questions correctly before you can proceed.

1. True or False? Sum 1 always has a larger value than Sum 2.
   ○ True.
   ◉ False.

2. True or False? Both sums can have the same value.
   ○ True.
   ◉ False.

3. True or False? You will only earn a bonus if you choose the option indicating the sum with the largest value.
   ◉ True.
   ○ False.

Check Answers

All answers are correct! Click 'Next' to proceed.

Next

Task 1 - Instructions

This task will have 32 rounds.
At the end of the task, one round will be chosen randomly and what you earned in that round will be added to your total earnings.
All rounds are equally likely to be selected.

For each round of this task, you will be shown a box with Blue and Red dots.
The box is equally likely to have more Blue or more Red dots.
There are always 2 more dots of one colour than the other.

In each round, you will be asked to choose between two options: Option A and Option B.
Each option pays £4.00 if it indicates the colour with more dots.

For example, you could be asked to choose between the following options:

Option A: **£4.00** if most dots are Red

Option B: **£4.00** if most dots are Blue

In order to get familiarised with the task, you will begin with a practice round. You will not be paid for the practice round.

Questions

You must answer the following questions correctly before you can proceed.

1. True or False? There are always more Red dots than Blue dots.
   - ○ True.
   - ◉ False.

2. True or False? There are always two more dots of one colour than another.
   - ◉ True.
   - ○ False.

3. True or False? You will only earn a bonus if you choose the option indicating the colour with more dots.
   - ◉ True.
   - ○ False.

Check Answers

All answers are correct! Click 'Next' to proceed.

Next

Lab Discounting.

Sums Task

Task 2 - Instructions

This task will have 32 rounds.
At the end of the task, one round will be chosen randomly and what you earned in that round will be added to your total earnings.
All rounds are equally likely to be selected.

For each round of this survey, you will be shown two sums, Sum 1 and Sum 2.
Each sum comprises many randomly selected integers ranging from 0 to 9.
Each of the sums is equally likely to have the largest value; there are no ties.

In each round, you will be asked to choose between two options: Option A and Option B.
Each option pays £4.00 if it indicates the sum with the largest value.

For example, you could be asked to choose between the following options:

Option A: **£4.00** if Sum 1 has the largest value

Option B: **£4.00** if Sum 2 has the largest value

In order to get familiarised with the task, you will begin with a practice round. You will not be paid for the practice round.

Recall that **your bonus for this task will be discounted by 3.0% for each second you take.** This means that, for example, if you take 10 seconds in this task, you would get about 26.25% less of whichever bonus you were to receive from this task.
Time spent on a particular task will not affect the bonus from other tasks.

Questions

You must answer the following questions correctly before you can proceed.

1. True or False? Sum 1 always has a larger value than Sum 2.
   ○ True.
   ◉ False.

2. True or False? Both sums can have the same value.
   ○ True.
   ◉ False.

3. True or False? You will only earn a bonus if you choose the option indicating the sum with the largest value.

- (•) True.
- ( ) False.

Check Answers

All answers are correct! Click 'Next' to proceed.

Next

Task 1 - Instructions

This task will have 32 rounds.
At the end of the task, one round will be chosen randomly and what you earned in that round will be added to your total earnings.
All rounds are equally likely to be selected.

For each round of this task, you will be shown a box with Blue and Red dots.
The box is equally likely to have more Blue or more Red dots.
There are always 2 more dots of one colour than the other.

In each round, you will be asked to choose between two options: Option A and Option B.
Each option pays £4.00 if it indicates the colour with more dots.

For example, you could be asked to choose between the following options:

Option A: **£4.00** if most dots are Red

Option B: **£4.00** if most dots are Blue

In order to get familiarised with the task, you will begin with a practice round. You will not be paid for the practice round.

Recall that **your bonus for this task will be discounted by 3.0% for each second you take.** This means that, for example, if you take 10 seconds in this task, you would get about 26.25% less of whichever bonus you were to receive from this task.
Time spent on a particular task will not affect the bonus from other tasks.

Questions

You must answer the following questions correctly before you can proceed.

1. True or False? There are always more Red dots than Blue dots.
   - ○ True.
   - ◉ False.

2. True or False? There are always two more dots of one colour than another.
   - ◉ True.
   - ○ False.

3. True or False? You will only earn a bonus if you choose the option indicating the colour with more dots.

- (●) True.
- ( ) False.

Check Answers

All answers are correct! Click 'Next' to proceed.

Next

Online Delay.

Sums Task

Task 1 - Instructions

This task will have 32 rounds.
At the end of the task, one round will be chosen randomly and what you earned in that round will be added to your total earnings.
All rounds are equally likely to be selected.

For each round of this survey, you will be shown two sums, Sum 1 and Sum 2.
Each sum comprises many randomly selected integers ranging from 0 to 9.
Each of the sums is equally likely to have the largest value; there are no ties.

In each round, you will be asked to choose between two options: Option A and Option B.
Each option pays £4.00 if it indicates the sum with the largest value.

For example, you could be asked to choose between the following options:

Option A: **£4.00** if Sum 1 has the largest value

Option B: **£4.00** if Sum 2 has the largest value

In order to get familiarised with the task, you will begin with a practice round. You will not be paid for the practice round.

You will need to wait 30 seconds before you can submit your answer in each round. This is done so that you take time understanding the task better.

Questions

You must answer the following questions correctly before you can proceed.

1. True or False? Sum 1 always has a larger value than Sum 2.
   - ○ True.
   - ◉ False.

2. True or False? Both sums can have the same value.
   - ○ True.
   - ◉ False.

3. True or False? You will only earn a bonus if you choose the option indicating the sum with the largest value.
   - ◉ True.

○ False.

Check Answers

All answers are correct! Click 'Next' to proceed.

Next

Task 2 - Instructions

This task will have 32 rounds.
At the end of the task, one round will be chosen randomly and what you earned in that round will be added to your total earnings.
All rounds are equally likely to be selected.

For each round of this task, you will be shown a box with Blue and Red dots.
The box is equally likely to have more Blue or more Red dots.
There are always 2 more dots of one colour than the other.

In each round, you will be asked to choose between two options: Option A and Option B.
Each option pays £4.00 if it indicates the colour with more dots.

For example, you could be asked to choose between the following options:

Option A: **£4.00** if most dots are Red

Option B: **£4.00** if most dots are Blue

In order to get familiarised with the task, you will begin with a practice round. You will not be paid for the practice round.

You will need to wait 30 seconds before you can submit your answer in each round. This is done so that you take time understanding the task better.

Questions

You must answer the following questions correctly before you can proceed.

1. True or False? There are always more Red dots than Blue dots.
   - ○ True.
   - ◉ False.

2. True or False? There are always two more dots of one colour than another.
   - ◉ True.
   - ○ False.

3. True or False? You will only earn a bonus if you choose the option indicating the colour with more dots.
   - ◉ True.

○ False.

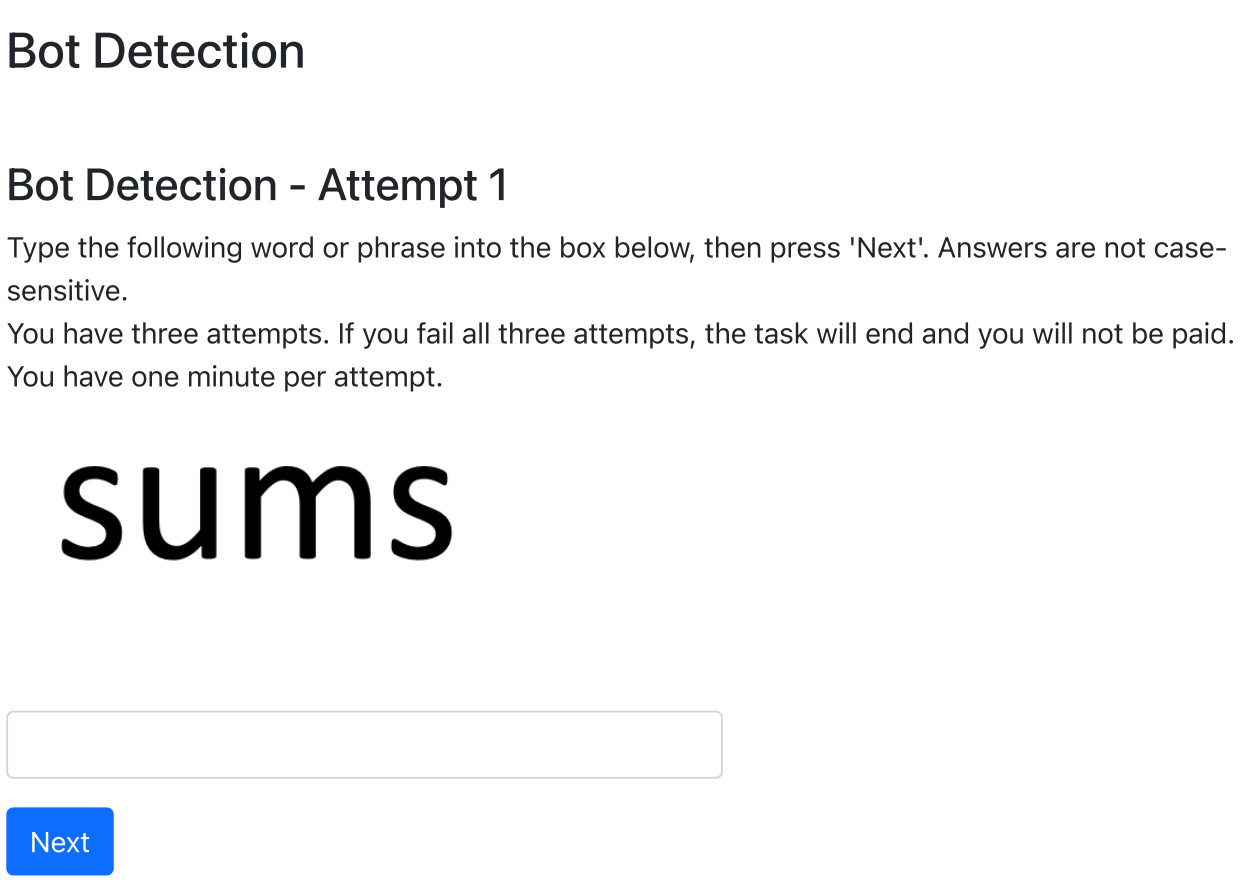
Check Answers

All answers are correct! Click 'Next' to proceed.

Next

Online Appendix B.3. Captcha

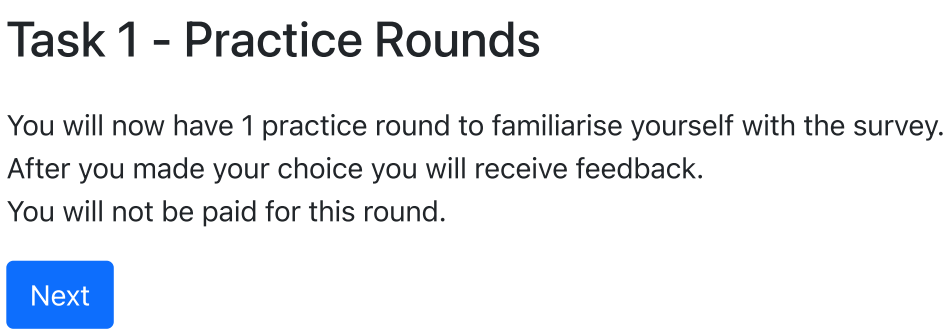
Online Baseline and Online Delay.



59

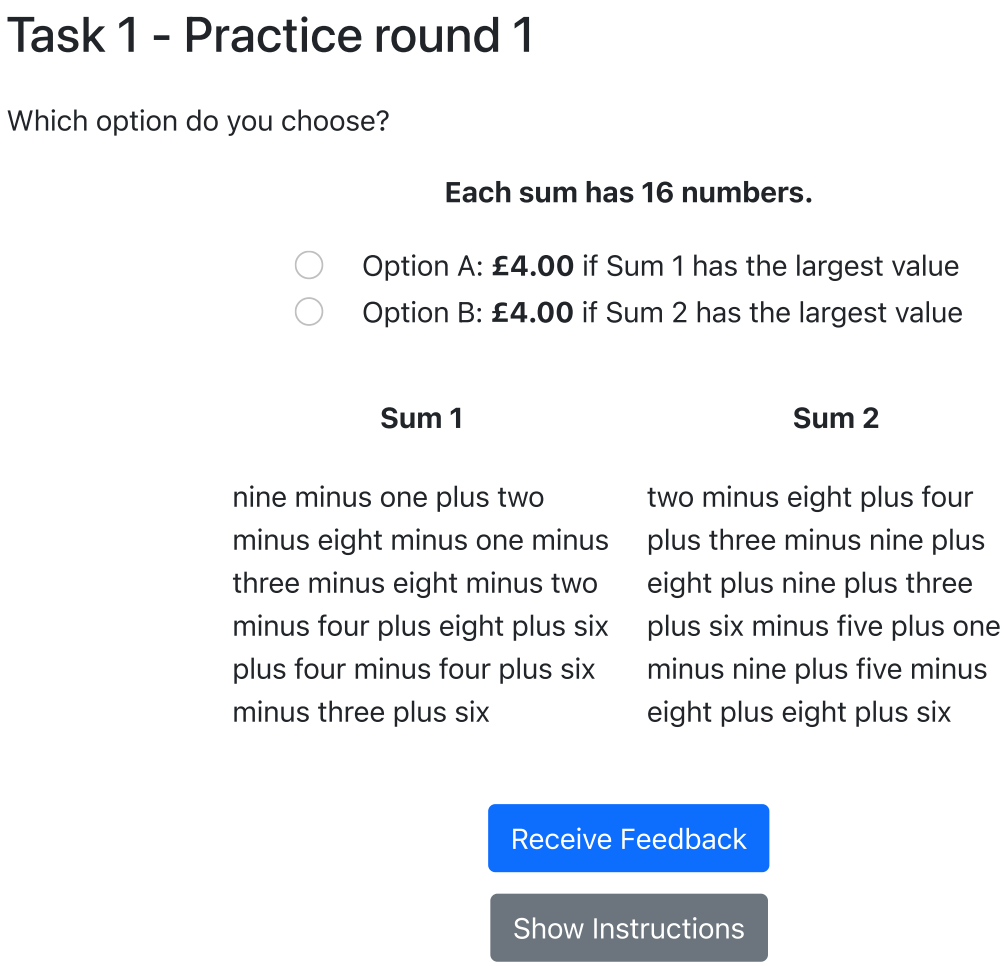
Online Appendix B.4. Practice Rounds

All Treatments.



Lab Baseline and Online Baseline.

Sums Task



Task 1 - Practice round 1

Which option do you choose?

Each sum has 16 numbers.

- (●) Option A: **£4.00** if Sum 1 has the largest value
- ( ) Option B: **£4.00** if Sum 2 has the largest value

Sum 1	Sum 2
nine minus one plus two	two minus eight plus four
minus eight minus one minus	plus three minus nine plus
three minus eight minus two	eight plus nine plus three
minus four plus eight plus six	plus six minus five plus one
plus four minus four plus six	minus nine plus five minus
minus three plus six	eight plus eight plus six

In this round, the sums added up to 7 and 16, respectively.
You chose Option A. If this were a paid round, you would have earned a bonus of 0.
If, instead, you had chosen Option B and this were a paid round, you would have earned a bonus of £4.00.

Next

Show Instructions

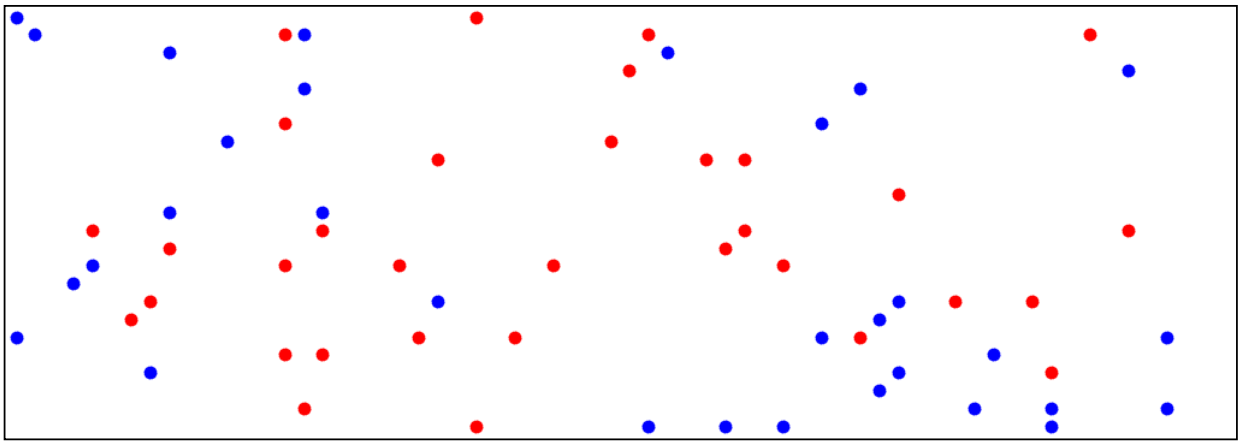
Dots Task

Task 1 - Practice round 1

Which option do you choose?

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

- Option A: **£4.00** if most dots are Blue
- Option B: **£4.00** if most dots are Red



Receive Feedback

Show Instructions

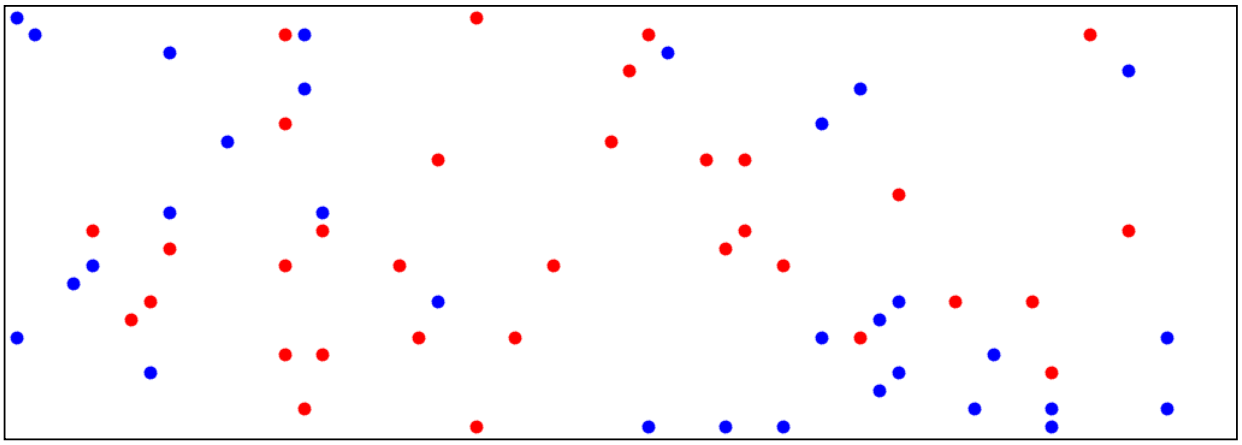
Dots Task Feedback

Task 1 - Practice round 1

Which option do you choose?

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

- (●) Option A: **£4.00** if most dots are Blue
- ( ) Option B: **£4.00** if most dots are Red



In this round there were 33 Red and 31 Blue dots.
You chose Option A. If this were a paid round, you would have earned a bonus of 0.
If, instead, you had chosen Option B and this were a paid round, you would have earned a bonus of £4.00.

Next

Show Instructions

Lab Discounting.

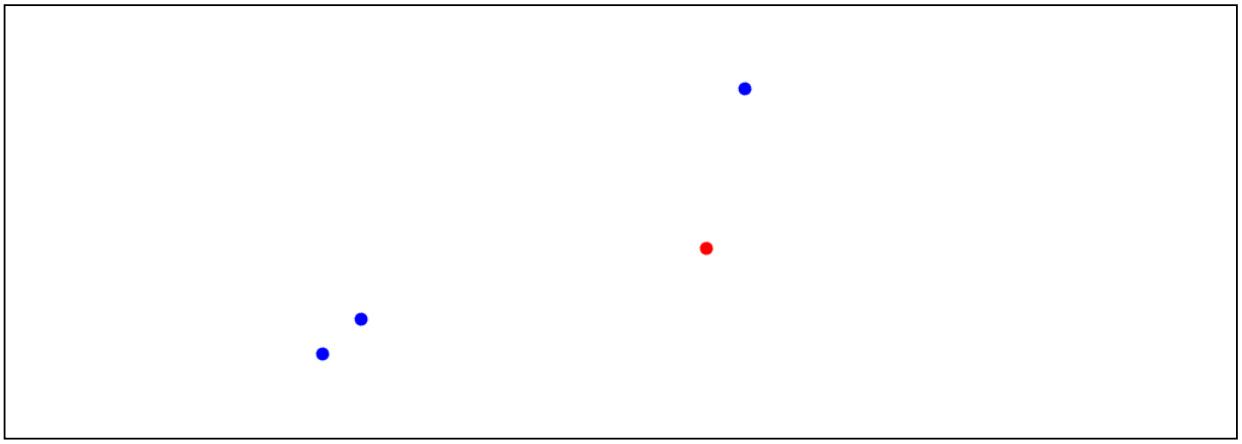
Sums Task

Task 1 - Practice round 1

Which option do you choose?

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

- Option A: **£3.79** if most dots are Blue
- Option B: **£3.79** if most dots are Red



Receive Feedback

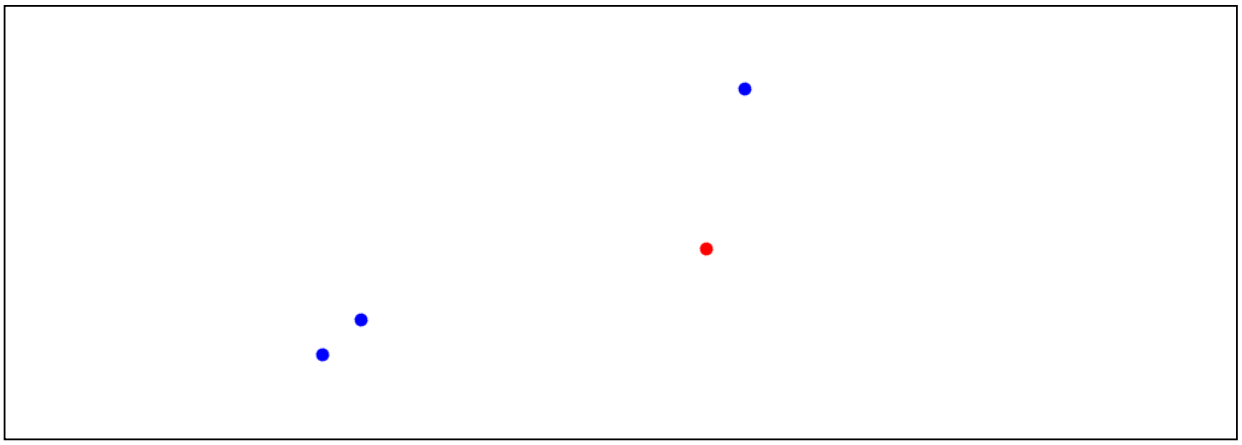
Show Instructions

Task 1 - Practice round 1

Which option do you choose?

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

- (●) Option A: **£1.64** if most dots are Blue
- ( ) Option B: **£1.64** if most dots are Red



In this round there were 1 Red and 3 Blue dots.
You chose Option A. If this were a paid round, you would have earned a bonus of £1.64.
If, instead, you had chosen Option B and this were a paid round, you would have earned a bonus of 0.

Next

Show Instructions

Task 2 - Practice round 1

Which option do you choose?

Each sum has 16 numbers.

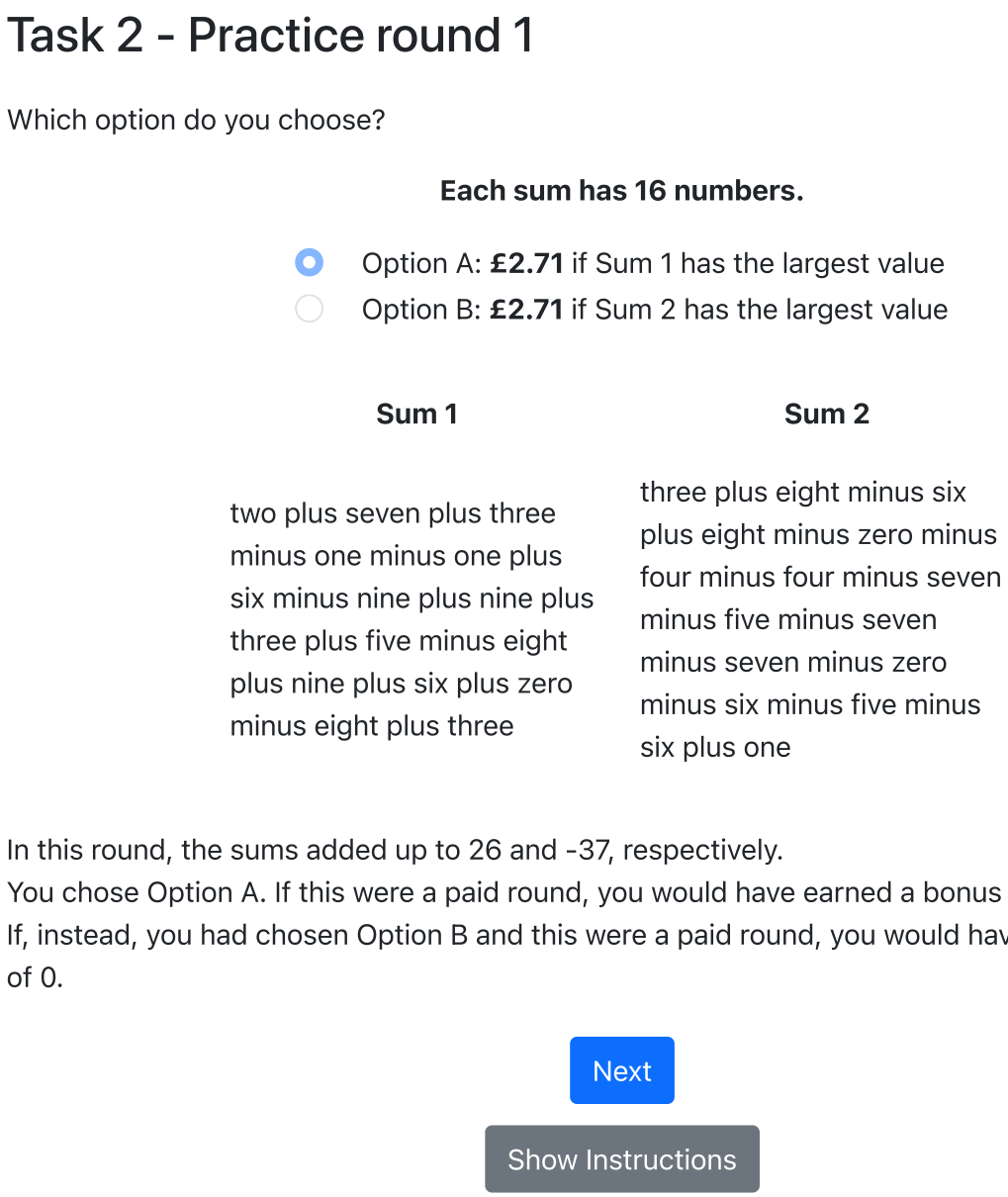
- ○ Option A: **£3.83** if Sum 1 has the largest value
- ○ Option B: **£3.83** if Sum 2 has the largest value

Sum 1	Sum 2
two plus seven plus three minus one minus one plus six minus nine plus nine plus three plus five minus eight plus nine plus six plus zero minus eight plus three	three plus eight minus six plus eight minus zero minus four minus four minus seven minus five minus seven minus seven minus zero minus six minus five minus six plus one

Receive Feedback

Show Instructions

Dots Task Feedback

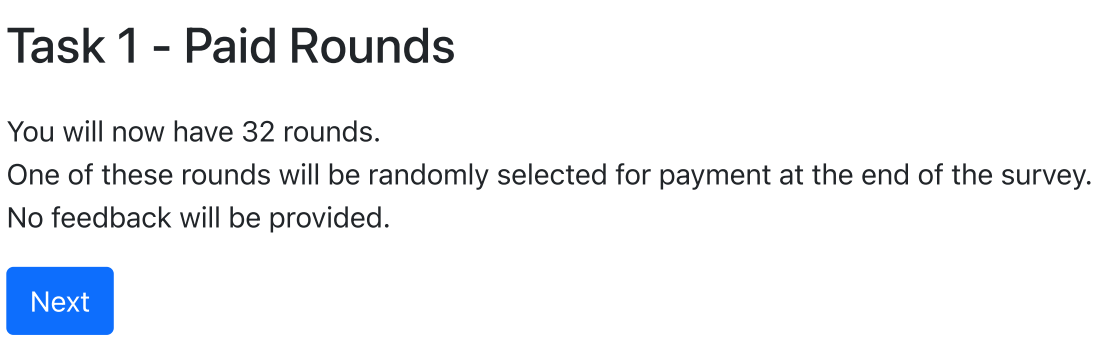


Online Delay.

Online Delay is identical to Lab Baseline and Online Baseline but for the fact that the 'Receive Feedback' button only appears after a 30 second delay.

Online Appendix B.5. Paid Rounds

All Treatments.



Lab Baseline and Online Baseline.

Sums Task

Task 1 - Round 1

Which option do you choose?

Each sum has 16 numbers.

- ○ Option A: **£4.00** if Sum 1 has the largest value
- ○ Option B: **£4.00** if Sum 2 has the largest value

Sum 1	Sum 2
two plus nine plus six minus seven minus seven minus nine plus six plus seven plus five minus zero minus seven plus nine plus seven plus five plus eight plus one	nine plus three plus zero minus six plus zero minus zero plus two plus zero minus zero minus six plus two plus three minus five minus three plus four minus seven

Next

Show Instructions

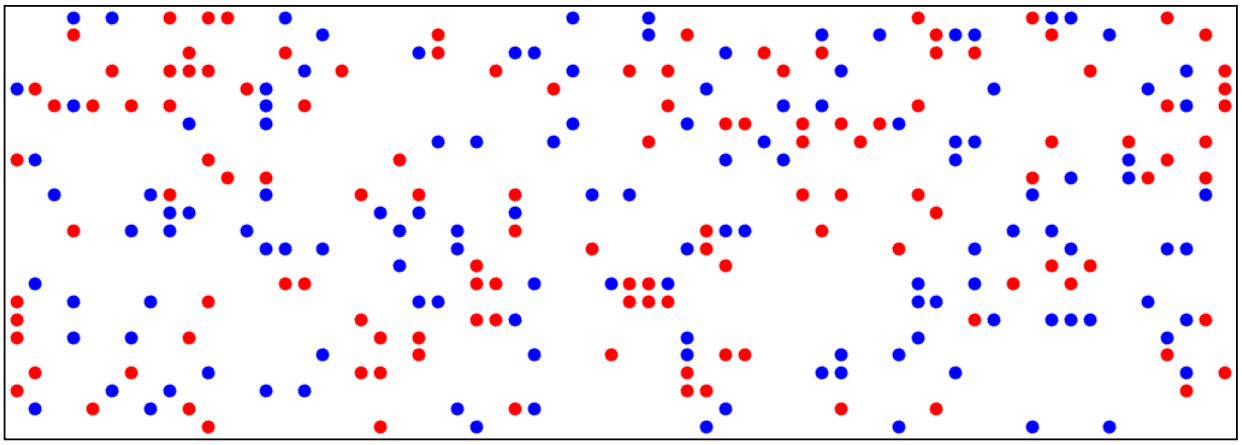
Dots Task

Task 1 - Round 1

Which option do you choose?

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

- (●) Option A: **£4.00** if most dots are Red
- ( ) Option B: **£4.00** if most dots are Blue



Next

Show Instructions

Lab Discounting.

Sums Task

Task 2 - Round 1

Which option do you choose?

Each sum has 16 numbers.

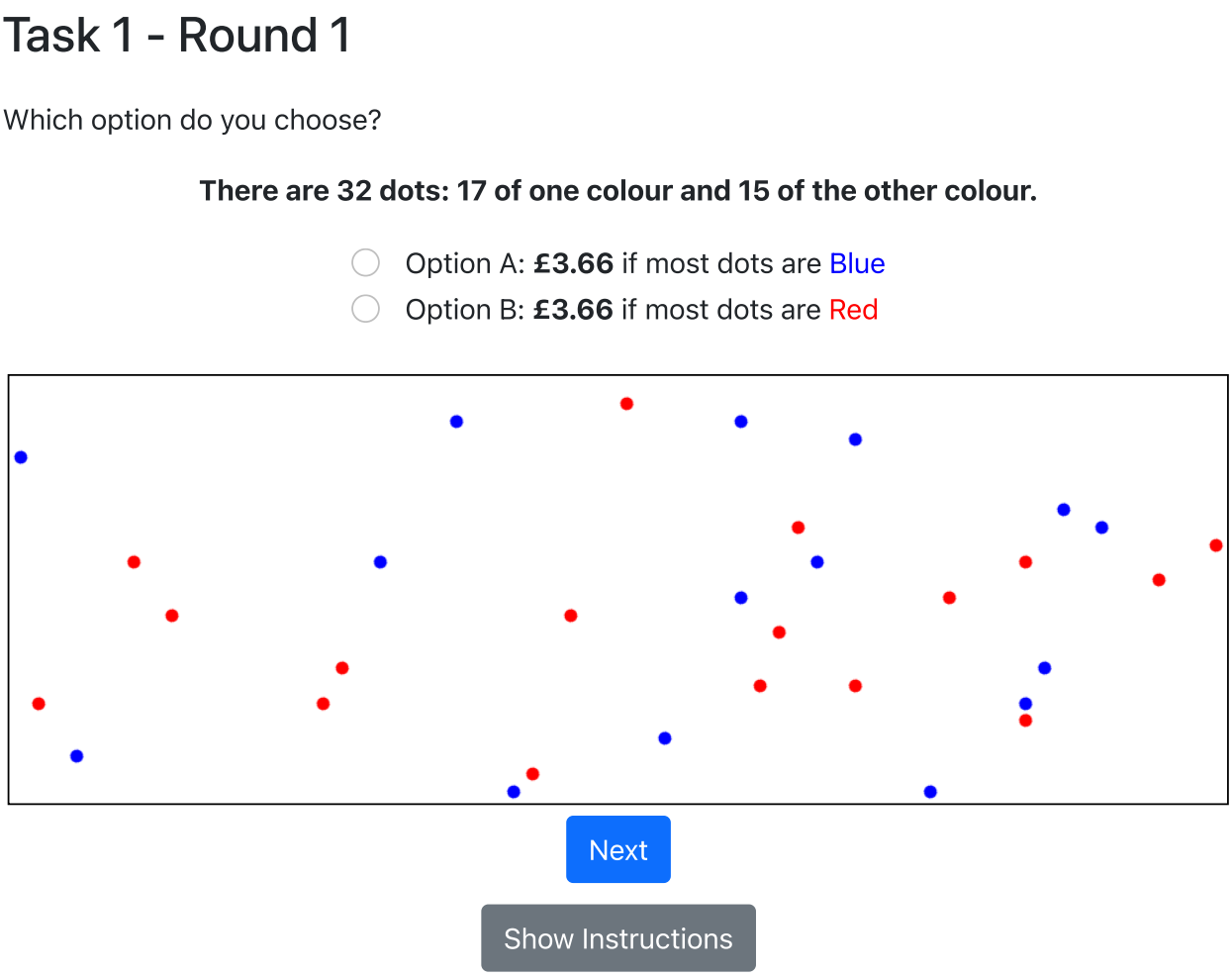
- ○ Option A: **£3.81** if Sum 1 has the largest value
- ○ Option B: **£3.81** if Sum 2 has the largest value

Sum 1	Sum 2
seven minus nine minus five minus three plus two plus five plus two minus nine minus three plus six minus nine plus zero plus six minus five minus eight plus five	one minus two plus seven minus eight minus zero minus seven minus eight plus seven plus eight minus six minus seven minus eight minus nine plus five minus seven minus three

Next

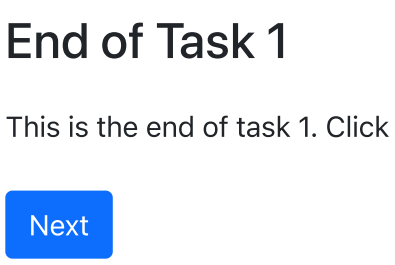
Show Instructions

Dots Task



Online Delay.
Online Delay is identical to Lab Baseline and Online Baseline but for the fact that the 'Next' button only appears after a 30 second delay.

All Treatments.



End of Task 2

This is the end of task 2. Click Next to continue to the next task.

Next