Online Appendix for

"Sequential Sampling by Individuals and Groups: An Experimental Study"

Pëllumb Reshidi* Alessandro Lizzeri† Leeat Yariv‡ Jimmy Chan§ Wing Suen¶

February 3, 2024

Abstract

In this Online Appendix, we describe the details of the experimental interface and display sample instructions. In addition, we offer several extensions to the underlying theoretical model and our data analysis. In particular, we illustrate the inconclusive effects of risk aversion and show that our risk and altruism elicitations have little explanatory power in our data. We inspect session effects and demonstrate that they are unlikely to generate our results. We also consider various levels of clustering in our analyses and offer supplementary analyses to those discussed in the text.

*Department of Economics, Princeton University preshidi@princeton.edu

†Department of Economics, Princeton University lizzeri@princeton.edu

‡Department of Economics, Princeton University lyariv@princeton.edu

§Department of Economics, National Taiwan University and Department of Economics, The Chinese University of Hong Kong jimmyhingchan@ntu.edu.tw

¶Faculty of Business and Economics, University of Hong Kong wsuen@econ.hku.hk

1 Lab Implementation

In what follows, we first describe several of our design choices and the workings of the experimental interface. We then offer sample instructions for our majority treatment.

1.1 Conveying Information

Given the information that unfolds (the evolution of the Brownian motion), we compute at every point in time the probability that choice A or choice B is correct and show this computation directly to participants. In doing so, we ensure that probabilities are adequately updated, and thus, none of our findings emerges as a direct consequence of participants' failure to compute Bayesian posteriors.

Figure 1: Information Bar

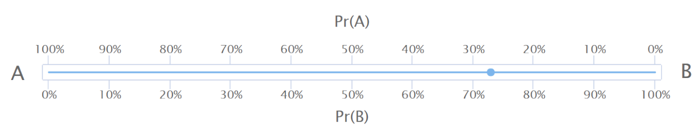
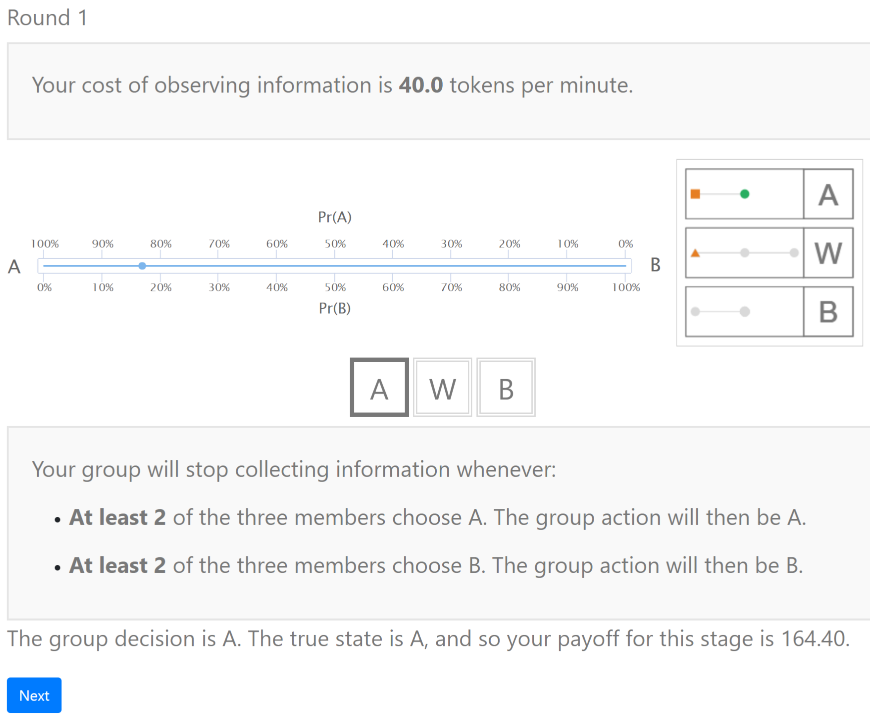


Figure 1 depicts the information bar through which participants are informed about the probability of choice A or choice B being correct. At the top, we depict the probability of choice A being correct, whereas, at the bottom, we depict $100 - P(A)$, or the probability that choice B is correct. At the beginning of each game, the blue dot (which in the figure is at 27% for A, or equivalently at 73% for B) is positioned exactly in the middle, indicating that initially the two choices are equally likely to be correct. As the Brownian motion evolves (which represents the log-likelihood of each state being correct), we transform it into a probability of choice A or choice B being correct. Namely we compute $P(A\,\text{is correct}) = \frac{e^{X_t}}{1+e^{X_t}}$ and accordingly position the blue dot.

1.2 Experimental Interface

The interface seen by participants in our majority treatment is shown in Figure 2.

Figure 2: Majority Treatment Interface



- On the top left corner of the interface there is a round counter. This ranges from Round 1 all the way to Round 30.

- Below the round counter, throughout the experiment, participants are reminded of their waiting costs—their information acquisition costs. These costs are the same for all treatments, namely 40 tokens per minute.

- Below these reported costs, participants see the information bar described in Section 1.1.

- To the right of the information bar, participants have access to a panel that informs them of the decisions of other group members. Participants always see their own position as the green circle, and the choices of other group members as the orange square and triangle. As can be seen in Figure 2, both the participant as well as another group member have voted for A. An analogous panel appears for treatments involving groups using unanimity; it is absent in our individual treatments.

- Beneath the information bar, participants see an A (vote for A), W(wait), and B (vote for B) buttons. By clicking on these buttons, participants can submit/change their votes. Each round starts with the W button as the default choice. The current choice is highlighted using a gray frame around the corresponding button. In Figure 2, the participant has clicked on A.

- Beneath the voting buttons, participants are once more reminded of the voting rule.

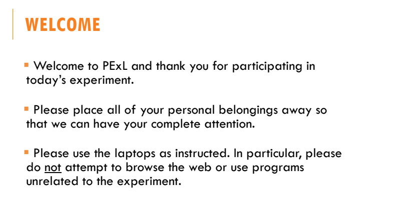
- Beneath the voting rule reminder, a new line appears after the pivotal vote is cast, informing the participants of the realized outcome, as well as their payoff. In the round depicted in Figure 2, the majority of participants chose option A, which matched the realized state, and their payoff was $200 - t \cdot 40 = 164.40$, where t represents the time group members took to arrive at the decision. Analogous reports occur for treatments in which groups use unanimity, or when individuals have full discretion.

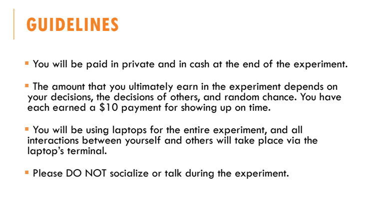
- Whenever participants are ready, they can indicate their desire to start the new round by clicking the "Next" button. Once all participants within the session are ready, new random groups are formed and the new round begins.

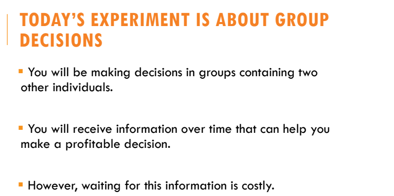
1.3 Sample Instructions

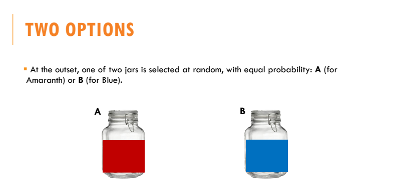
1.3.1 Initial Instructions

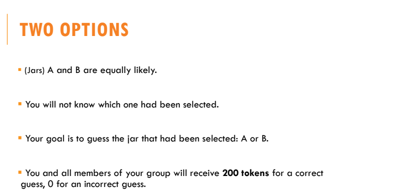
The initial instructions are identical for all treatments. Each treatment started with the instructions being read aloud, as well as two practice round for the participants to get used to the interface.

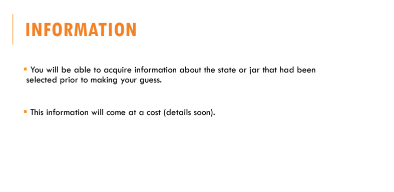


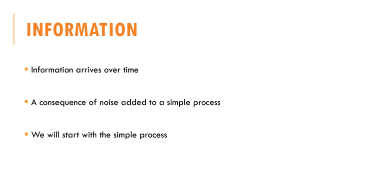


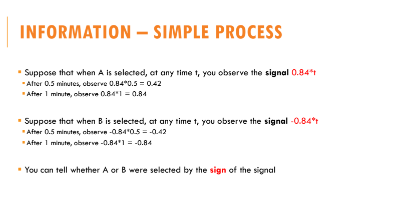


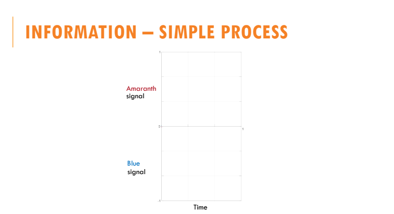


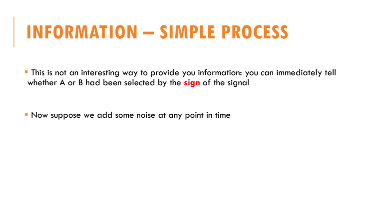


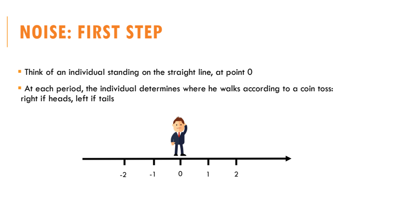


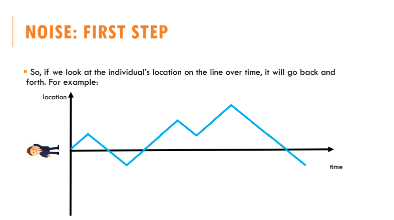


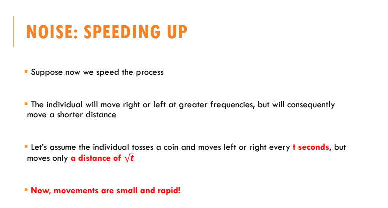


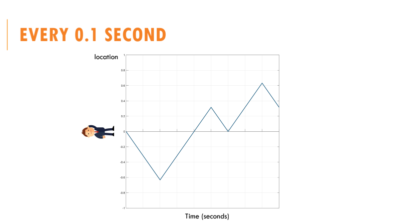


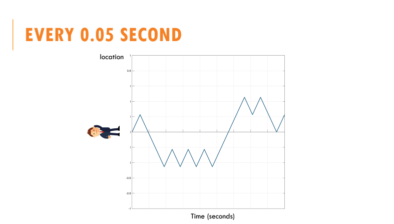


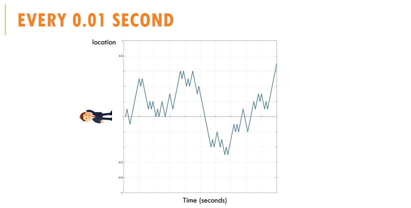


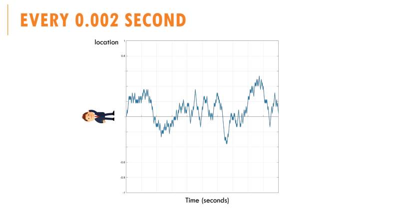


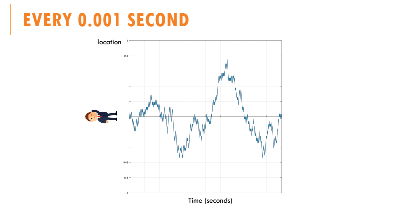


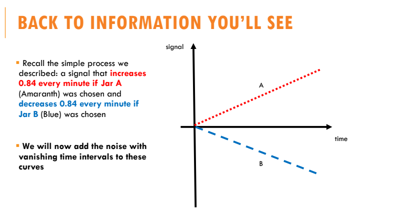


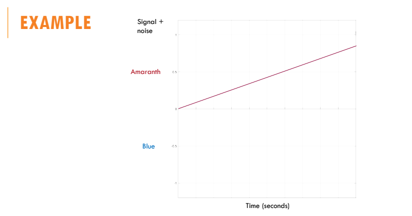


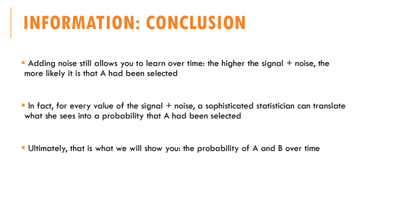






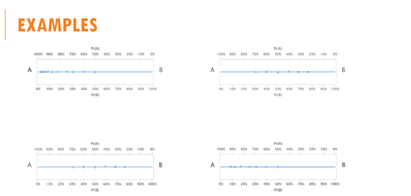


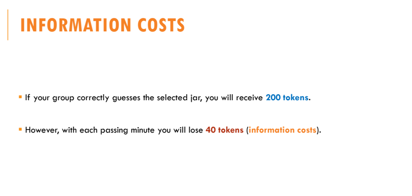


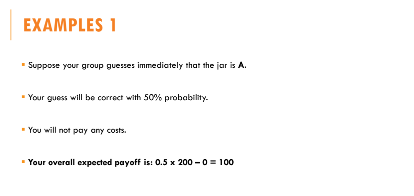


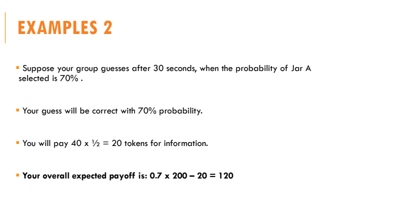
1.3.2 Sample Instructions: Majority Treatment

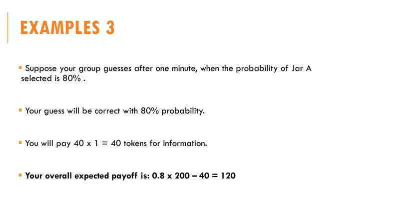
The examples of the process evolving were animated.

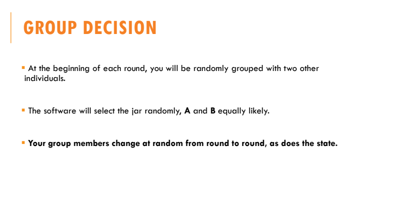


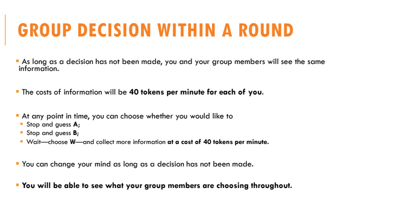


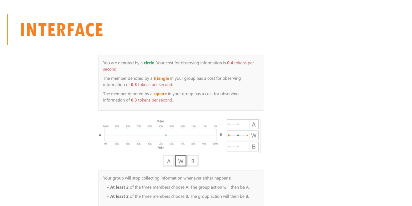


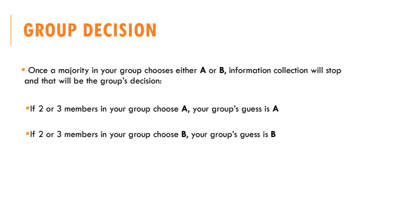


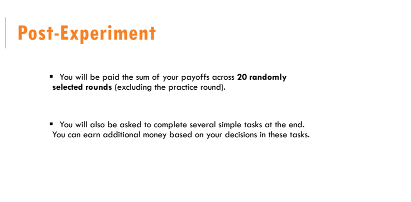












2 Theoretical Predictions

We now outline the theoretical predictions for our various treatments. For details, see Dvoretzky et al. (1953) or Chan et al. (2018).

We consider the setting described in our experimental design. An agent assesses which one of two ex-ante equally likely states, A or B, are realized. Information follows a Wiener process with a variance of 1. When the state is A, the process has drift $\mu = 0.84$; When the state is B, the process has drift $-\mu = -0.84$. Tracking this information comes at a flow cost of c. The agent guesses the state that is more likely once information collection terminates. For ease of exposition, we normalize the reward for an ultimately correct guess of the state to be 1. With this normalization, the flow cost corresponding to that used in our experiments is $c = 0.2$.

It is convenient to define $\mu' \equiv 2\mu^2$. The agent's posterior belief is then given by a Wiener process, with drift μ' and instantaneous variance $2\mu'$ in state A, and drift $-\mu'$ and instantaneous variance $2\mu'$ in state B. For our parameters, $\mu' = 1.4$.

One of the main contributions of Wald (1945) and the continuous-time counterpart of Dvoretzky et al. (1953) is to demonstrate that, in the sequential-sampling setting, an optimizing agent uses a simple threshold policy. Namely, at any time t, the agent calculates the log-likelihood ratio $\theta_t = \log\left(\frac{\Pr[A]}{\Pr[B]}\right)$. The optimal policy specifies a pair of cutoffs (g, G), with $G \geq g$, such that the agent stops information collection and guesses the state is A whenever $\theta_t \geq G$. Similarly, the agent stops information collection and guesses the state is B whenever $\theta_t \leq g$.

For $\theta \in [g, G]$, let $u(\theta|g, G)$ represent the expected payoff from the deliberation process. A similar derivation to that of Chan et al. (2018) yields[1]

$$\begin{aligned} u(\theta|g,G) = {} & \frac{e^G(e^\theta - e^g) + (e^G - e^\theta)}{(1+e^\theta)(e^G - e^g)} \\ & - \frac{c}{\mu'} \frac{(G-\theta)(e^{G+\theta} + e^g) + (\theta - g)(e^{g+\theta} + e^G) - (G-g)(e^\theta + e^{G+g})}{(1+e^\theta)(e^G - e^g)}. \end{aligned}$$

The corresponding first-order condition with respect to the lower boundary is then[2]

$$\frac{\partial u(\theta|g,G)}{\partial g} = \frac{-(e^G - e^\theta)}{(1+e^\theta)(e^G-e^g)^2}\left[e^g(e^G - 1) - \frac{c}{\mu'}\left((G-g)e^g(e^G-1) + (e^G - e^g)(1-e^g)\right)\right] = 0.$$

[1]Our formulation here differs from that of Chan et al. (2018) in that they consider exponentially discounted utilities, whereas we consider flow costs of time spent on information collection. This modification simplifies the experimental interface.

[2]The first-order approach is indeed valid, we omit details for the sake of brevity.

This condition shows that the cutoffs satisfying the first-order condition do not depend on the current log-likelihood ratio θ. Thus, solutions are stationary.

Because the problem is symmetric, the solution satisfies $g = -G$. The optimal value of G can then be determined by the implicit function $c\left(2e^{G}G + e^{2G} - 1\right) - e^{G}\mu' = 0$. With $\mu' = 1.4$ and $c = 0.2$, the numerical solution for the optimal boundary is $G^* = 1.46$. Translated into probabilities, this value becomes $\frac{e^{1.46}}{1+e^{1.46}} = 0.81$. Thus, a risk-neutral agent should wait until the probability of the most likely state is 81%. Once again, our parameter choices ensure that the optimal stopping threshold is not overly extreme, so that errors are unlikely to be one-sided.

Consider now a group of $n > 1$ homogeneous agents. At each point in time, each agent decides whether she would like to stop and guess A, stop and guess B, or wait. The group continues information collection until either a majority or a unanimity of agents in the group choose to guess the same state.

The utilitarian efficient equilibrium for the group, under both majority and unanimity, corresponds to the optimal search policy described above, namely utilizing a threshold of 81%. In our experiment, participants make group decisions with other individuals who do not choose the exact 81% threshold. However, it is easy to show that, as long as agents use symmetric cutoff policies, the 81% threshold is still the best response for any agent. Therefore, there are no immediate consequences for optimal choices because of potentially noisy behavior of other agents.

3 Beyond Risk Neutrality

Consider the individual case. Suppose an agent uses a threshold posterior of $\tilde{p}$. This threshold gives rise to a distribution of end times, $f(t|\tilde{p})$ (for which only a Fourier series representation can be constructed). For any $\hat{t}$ at which information-collection terminates, the agent receives the following lottery

$$\tilde{p}u(x - c\hat{t}) + (1 - \tilde{p})u(-c\hat{t}).$$

The agent would be choosing the optimal $\tilde{p}$ to maximize her expected utility

$$\max_{\tilde{p}\in[0.5,1]} \int_0^\infty \left(\tilde{p}u(x - cs) + (1 - \tilde{p})u(-cs)\right) f(s|\tilde{p})ds.$$

By choosing a larger $\tilde{p}$ the agent minimizes the uncertainty in the lottery she receives. However, this increases the uncertainty regarding the time it takes to reach a decision. The effects of risk are, again, unclear.

Below, we use our risk elicitations to illustrates that risk, indeed, has limited explanatory power in our data.

4 Additional Analysis

4.1 Pulling the Trigger: Majority and Unanimity

In the main text, we depict the evolution of posteriors and the corresponding choices in our individual treatments. Figure 3 and Figure 4 below provide analogous graphs for our majority and unanimity treatments, which depict decision posteriors corresponding to pivotal votes.

The general patterns observed for our individual treatments remain. For example, later decisions in our dynamic treatments often correspond to lower accuracies. However, there are some differences. In particular, groups using majority pull the trigger far quicker than groups using unanimity, in line with results described in the text.

Figure 3: Pulling the Trigger: Majority Treatment

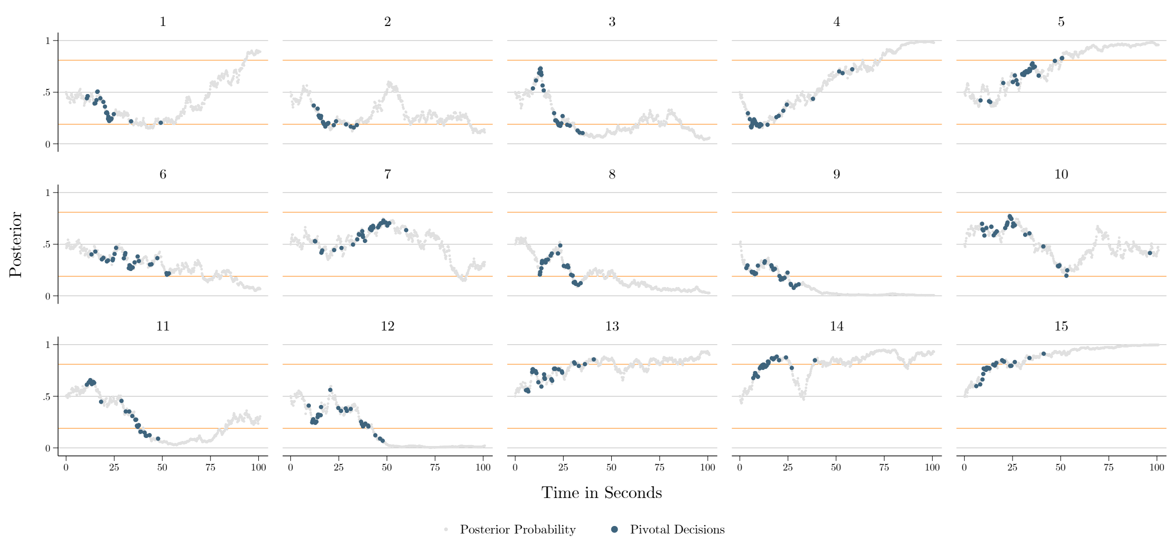
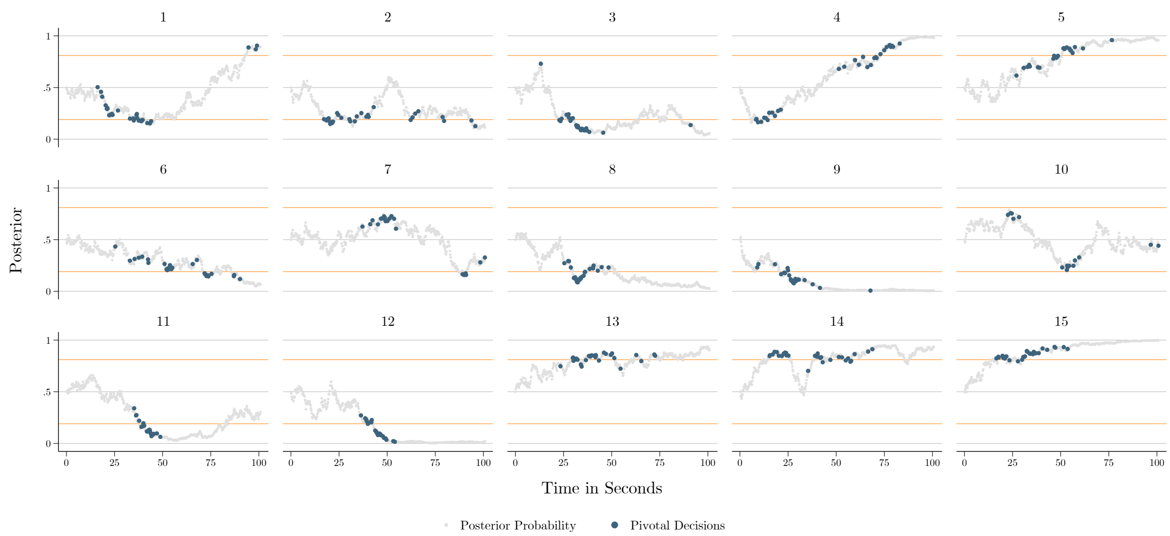


Figure 4: Pulling the Trigger: Unanimity Treatment



4.2 Non-stationary Threshold Posteriors

In the text, we consider individuals' voting probabilities in each of our treatments. We now take a different approach, analyzing the realized decision posteriors when the pivotal vote is cast. Table 1 displays regression analysis pertaining to individual and group choices—the stopping posterior—in our treatments. We use the shorthand of I, M, and U for the individual, majority, and unanimity treatments, respectively. The variables d_M and d_U are dummy variables for the majority and unanimity treatments. To allow for learning, we include dummy variables of the form *Last 15 X*, with X denoting the treatment; these indicate whether observations are taken from the last 15 rounds of our sessions. Last, we consider the impacts of time spent collecting information to potentially account for non-stationary thresholds. We do so in two ways. First, we classify the processes as "Slow" or "Quick". For this classification, we calculate the time it takes to reach the theoretically optimal threshold of 0.81 in each process. If a process takes more time than the median process to pass the 0.81 threshold (i.e., 29.8 seconds) we label it "Slow"; otherwise, the process is labeled "Quick". The resulting variable *Slow X* is a dummy variable indicating whether a process is slow in each treatment X. We also consider the time spent collecting information in each treatment X, denoted by *Time X*. The last three specifications allow for fixed effects corresponding to the individuals casting the pivotal votes. Errors are clustered at the individual level.

The first column of Table 1 echoes our observations in the text. We see significant differences

Table 1: Decreasing Thresholds

	Posterior				
	Ordinary Regression			Fixed Effects Regression	
	All Rounds		Last 15 Rounds	All Rounds	Last 15 Rounds
Constant	0.755***	0.785***	0.806***		
	(0.00846)	(0.00738)	(0.0109)		
d_M	-0.0362***	-0.0303***	-0.0372***		
	(0.0112)	(0.0107)	(0.0128)		
d_U	0.0444***	0.0347***	0.0431***		
	(0.0103)	(0.00885)	(0.0124)		
Last 15 I	0.0247***	0.0247***		0.0299***	
	(0.00647)	(0.00647)		(0.00790)	
Last 15 M	0.0162***	0.0162***		0.0224***	
	(0.00613)	(0.00611)		(0.00653)	
Last 15 U	0.0376***	0.0376***		0.0430***	
	(0.00717)	(0.00688)		(0.00726)	
Slow I		-0.0648***	-0.0576***		
		(0.00557)	(0.00625)		
Slow M		-0.0774***	-0.0736***		
		(0.00717)	(0.0101)		
Slow U		-0.0440***	-0.0271***		
		(0.00652)	(0.00989)		
Time I				-0.000651***	-0.00110***
				(0.000209)	(0.000238)
Time M				-0.00130***	-0.00165***
				(0.000340)	(0.000523)
Time U				-0.000524***	-0.000723***
				(0.000132)	(0.000218)
N	1980	1980	990	1980	990

Standard errors in parentheses
Individual-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

between treatments, with less precise, or hasty, majority decisions and more precise, or slower, unanimous decisions. Compared to the individual treatment, the mean posterior with which the pivotal majority vote is cast is about 4 percentage points lower, whereas the mean posterior with which the pivotal vote is cast is about 4 percentage points higher.

Throughout, we see a significant effect of learning over the first 15 rounds, with participants becoming more patient, casting their vote with a significantly higher decision posterior. Because both the individual and majority treatments yield, on average, posteriors well below the theoretically optimal, the increase in decision posteriors in later rounds is a move towards the optimal choice. In the unanimity treatment, however, learning leads to overshooting, with an average decision posterior of 0.84 in the last 15 rounds. As mentioned at the outset, and elaborated on below, we do not see evidence of substantial learning beyond the first 15 rounds.

The second and third columns consider the impacts of the underlying process, i.e., whether it is slow or quick. Slow processes are associated with significantly lower decision posteriors across all our dynamic treatments. This association is present and similar in both magnitude and significance, even when restricting attention only to the last 15 rounds of each session. It is most pronounced for groups deciding through majority rule, and least pronounced in groups using unanimity. Lower decision posteriors in slow processes indicate a non-stationary threshold for halting information collection. The last two columns of Table 1 illustrate a declining-threshold pattern more directly, and echo the results presented in the text. Namely, we introduce an explicit dependence on the time at which a pivotal vote is cast.[3] The estimated coefficients corresponding to decision times are negative and statistically significant: the longer it takes for the pivotal vote to be cast, the lower is the threshold posterior. As before, the least affected treatment is unanimity and the most affected treatment is majority. In particular, in the majority treatment, in the last 15 rounds, for each 5 seconds that the group decision is delayed, the average threshold posterior decreases by almost one percentage point.

4.3 Estimation with Implementation Trembles

We now consider the possibility that participants implement their optimal threshold with trembles.

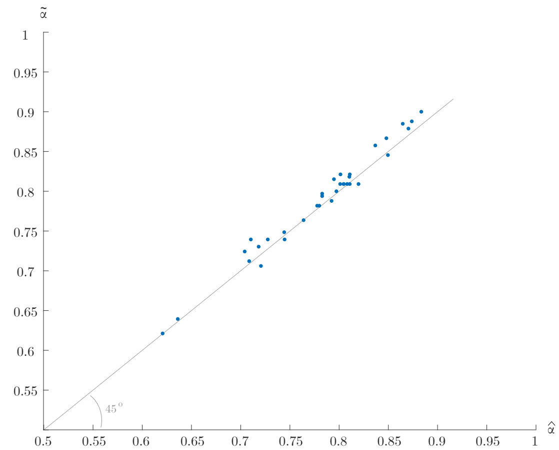
Suppose that instead of casting their vote based on their preferred threshold $f(t)$, participants cast their vote based on $f(t) + \varepsilon_t$, where ε_t is drawn from a normal distribution with mean 0 and standard deviation σ_ε.

In order to estimate the true $f(t)$, we first calculate, in our data, the average time participants take to cast a vote $\bar{t}$, and the observed estimated path of stopping posteriors—identified by an intercept $\hat{\alpha}$ and slope $\hat{\beta}$— derived from running an individual-level fixed-effects linear regression on the individual dynamic treatment data. In our estimation exercise, we match these three "moments" of our data: $\bar{t}$, $\hat{\alpha}$, and $\hat{\beta}$.

Specifically, we simulate Brownian paths with the parameters utilized in the experiment. We also simulate potential thresholds with different intercepts $\tilde{\alpha}$ and slopes $\tilde{\beta}$. Afterwards, we implement the decisions with different noise levels $\tilde{\sigma}_f$. For each conjectured $\{\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}_f\}$, we calculate the square distance from the the three moments we match, namely $\bar{t}$, $\hat{\alpha}$, and $\hat{\beta}$. Technically, we estimate our parameters of interest via the method of simulated moments (MSM). Our estimated

[3] The fixed-effects specification is appropriate since, without it, we could in principle identify a misleading positive association between decision times and decision posteriors. Indeed, mechanically, since we consider a diffusion with drift, posteriors exhibit an increasing trend. Group fixed effects cannot be used due to the random matching protocol we utilize. We therefore use pivotal-voter fixed effects to adequately capture the response to time passed.

Figure 5: Individual Intercepts With and Without Trembles



The graph above depicts the fixed-effects estimates of individual-level intercepts $\hat{\alpha}$ displayed on the horizontal axis and corrected estimates of individual level intercepts $\tilde{\alpha}$ on the vertical axis.

parameter values are then $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\sigma}_\varepsilon$ that minimize the sum of squared errors.[4]

Figure 5 plots the individual level intercepts estimated with and without trembles. From our observed data, we estimate the average intercept as $\hat{\alpha} = 0.782$, while the estimated slope is $\hat{\beta} = -0.000457$. After accounting for potential implementation trembles, we estimate the variance of the implementation error to be $\sigma_\varepsilon = 0.01$, $\alpha = 0.789$, and $\beta = -0.000486$. Thus, with this approach, and with these utilized moments, we do not find an economically significant difference from the estimations in which we simply ignore potential implementation trembles.

4.4 Learning within Sessions

In order to assess learning in our dynamic treatments, we examine whether there is a trend in participants' stopping posteriors over the course of our sessions. In Table 2 we regress participants' stopping posteriors on *Round*, which stands for the session round; *Slow*, which identifies the process occurring during the round as a slow or a quick process (see Section 4.2 above); and an interaction between *Round* and *Slow*, allowing for a different learning trend depending on the process.[5] We run an individual-level fixed-effects regression, allowing for a different intercept for each participant. By running the regression separately for each dynamic treatment, we allow for learning to affect

[4] We use Monte Carlo simulations to show that this method is indeed reliable in our setting, consistently estimating the true parameter values.

[5] We showed that participants tend to vote with a lower posterior when faced with a slow process, which is why we allow for different slopes and intercepts depending on the features of the process. Otherwise, if earlier rounds entail quicker processes, for example, we could erroneously infer a declining stopping posterior.

Table 2: Learning throughout Sessions

	Posterior								
	Individual Treatment			Majority Treatment			Unanimity Treatment		
	All Rounds	First 15	Last 15	All Rounds	First 15	Last 15	All Rounds	First 15	Last 15
Round	0.00154***	0.00511***	-0.000277	0.00150***	0.00177	0.00271***	0.00192***	0.00267***	0.00360***
	(0.000555)	(0.00114)	(0.00118)	(0.000420)	(0.00150)	(0.000742)	(0.000276)	(0.000788)	(0.000764)
Round × *Slow*	0.000619	0.00223	0.00190	-0.000186	0.0110***	0.00184	0.000912*	0.00904***	-0.00246*
	(0.000445)	(0.00249)	(0.00163)	(0.000701)	(0.00253)	(0.00175)	(0.000515)	(0.00182)	(0.00134)
Slow	-0.0705***	-0.0860***	-0.101**	-0.0712***	-0.168***	-0.119***	-0.0782***	-0.147***	-0.000178
	(0.00882)	(0.0220)	(0.0372)	(0.0133)	(0.0251)	(0.0405)	(0.00966)	(0.0169)	(0.0296)
$Correct_{t-1}$	-0.0218***	-0.0402***	-0.00913	-0.0256***	-0.0333***	-0.0183*	-0.0287***	-0.0247***	-0.0364***
	(0.00655)	(0.00910)	(0.00997)	(0.00687)	(0.00912)	(0.00969)	(0.00588)	(0.00643)	(0.00968)
$Difference_{t-1}$				0.0367	0.0290	-0.0172	0.0417	0.0319	-0.0348
				(0.0371)	(0.0610)	(0.0427)	(0.0282)	(0.0338)	(0.0387)
Individual Level FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	986	476	510	728	339	389	1392	672	720

Standard errors in parentheses
Individual-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

these treatments differently. To see whether there were enough rounds for learning to converge, we run additional regressions separately for the first and the last 15 rounds. In addition, we control for $Correct_{t-1}$ that equals 1 if the previous period's individual decision, or group decision in the majority and unanimity treatment, was correct, and equals 0 if the previous period's decision was incorrect. Finally, we control for $Difference_{t-1}$, which equals the difference between participants' last-period choice from the mean stopping posterior of other members of their group in the last period (for our majority and unanimity treatments only).

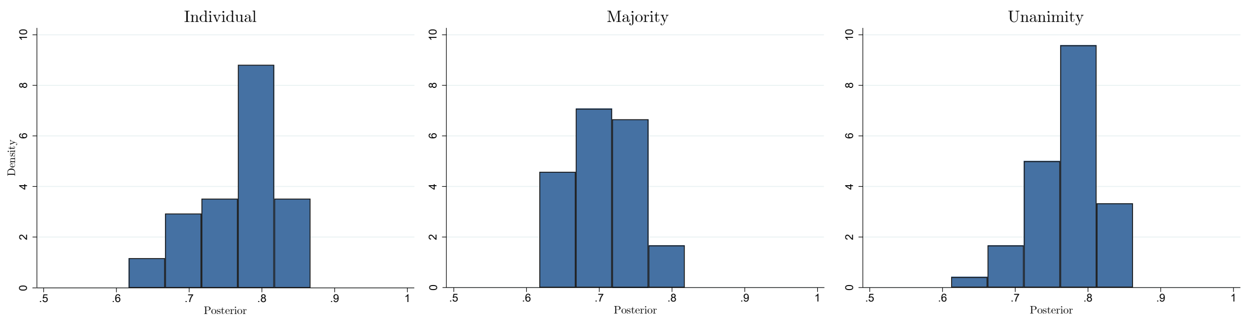
From the estimated coefficients of $Correct_{t-1}$, we see that, on average, participants cast their individual votes with a lower posterior in round t if their or their group's guess in round $t-1$ was correct. In contrast, the coefficients of $Difference_{t-1}$ is never statistically significant, implying that group effects operate more forcefully through the outcomes they generate.

Importantly, when it comes to learning, the regressions in the second and third columns reveal that both the magnitude and statistical significance of *Round* and *Round* × *Slow* drop in the last 15 rounds in the individual treatment. A similar decrease is observed for our majority and unanimity treatments. Even in cases where statistical significance persists, the magnitude is much lower in the last 15 round. The finding that the magnitude of learning is substantially lower in the last 15 rounds compared to the first 15 rounds, as well as the decrease in statistical significance, leads us to believe that 30 rounds afforded sufficient learning opportunities.

4.5 Individual-Level Choice Heterogeneity

Figure 6 depicts the distributions of the mean posteriors at the time of voting in our treatments, where mean posteriors are calculated for each participant separately.

Figure 6: Dynamic Treatments Individual Heterogeneity



In line with our discussion in the text, in the individual and unanimity treatments, a substantial fraction of individuals consistently vote at a posterior close to the optimal threshold. Under majority, we see more variation across individuals.

4.6 Observed and Simulated Treatment Groups

In addition to the cumulative distribution plots and the Kolmogorov-Smirnov tests appearing in the main text, below we present regressions in which we estimate the mean posterior of the observed and simulated group treatments. Concretely, in Table 3, d_{Sim} is a dummy variable equal to 0 for observed data points, and 1 for simulated data points. The constant captures the mean posterior in the observed data, whereas the coefficient of d_{Sim} captures the difference in mean posteriors between the observed and simulated data. We cluster errors at the individual level.

Table 3: Observed and Simulated Groups

	Posterior	
	Majority	Unanimity
d_{Sim}	0.0477***	0.000878
	(0.00847)	(0.00756)
Constant	0.727***	0.818***
	(0.00678)	(0.00454)
N	330480	330480

Standard errors in parentheses
Individual-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In line with the conclusions drawn in the main text, decision posteriors in the majority treatment are significantly lower than those derived from simulated groups of individuals using majority rule. In contrast, decision posteriors in the unanimity treatment are not significantly different than those derived from simulated groups of individuals using unanimity.

4.7 Voting Probabilities in the Last 15 Rounds

Table 4 replicates the analysis of individual voting probabilities reported in the text, restricting attention to the last 15 rounds of sessions. The results are qualitatively similar to those pertaining to data from all rounds, albeit less significant due to the reduction in power.

Table 4: Probit Regression: Last 15 Rounds

	$P(Vote)$					
	Individual	Majority	Unanimity	Individual	Majority	Unanimity
Posterior	5.421***	5.057***	5.779***	5.541***	3.826***	5.793***
	(0.524)	(0.495)	(0.520)	(0.601)	(0.605)	(0.557)
Time	0.218*	0.711***	0.396***	0.330***	0.611***	0.410***
	(0.128)	(0.179)	(0.121)	(0.122)	(0.191)	(0.118)
Slope				0.123**	0.107**	0.0223
				(0.0541)	(0.0463)	(0.0373)
Standard Dev				-0.265	0.989***	-0.122
				(0.437)	(0.379)	(0.345)
Constant	-5.138***	-4.604***	-5.507***	-5.361***	-4.007***	-5.521***
	(0.464)	(0.342)	(0.424)	(0.531)	(0.451)	(0.466)
N	4335	3553	6201	3810	2822	5474

Standard errors in parentheses
Individual-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.8 Risk Aversion, Altruism, and Alternative Clustering

In this section, we analyze alternative specifications for the analysis presented in the main text. The first column in Table 5 reports results from regressions focused on our main treatment effects in which standard errors are clustered at the individual level. The regressions include controls for risk attitudes and altruism, through two new explanatory variables: *Tokens Sent* and *Tokens Not Invested*. As mentioned in our description of the experimental design, at the end of each session, participants completed two risk-elicitation tasks as in Gneezy and Potters (1997). Namely, participants had 200 tokens to invest in a safe or risky asset. Tokens that were not invested were kept in the safe asset. The variable *Tokens Not Invested*, which can take values between 0 and 200, represents the amount participants decided to keep in the safe asset (and not invest in the risky asset).[6] Roughly speaking, the higher this value, the more risk averse participants are. At the end of each session, participants also played a dictator game, in which they were given 200 tokens and decided how much to keep for themselves, and how much to give to another, randomly-paired participant. The variable *Tokens Sent* represents the amount of tokens participants gave.[7] Since we elicit each measure twice, we can run an instrumental-variable regression, using the first elicitation as an instrument for the second. Doing so accounts for the fact that these are noisy elicitations, see Gillen et al. (2019).

Table 5: Alternative Specifications

	Posterior						
	Individual Level Clustering	No Clustering			Process Level Level Clustering		
	All Rounds	All Rounds		Last 15 Rounds	All Rounds		Last 15 Rounds
Constant	0.744***	0.767***	0.755***	0.744***	0.767***	0.755***	0.744***
	(0.0371)	(0.00293)	(0.00411)	(0.00993)	(0.0119)	(0.0135)	(0.0123)
d_M	-0.0329**	-0.0404***	-0.0362***	-0.0329***	-0.0404***	-0.0362***	-0.0329***
	(0.0134)	(0.00519)	(0.00727)	(0.00738)	(0.00539)	(0.00688)	(0.00640)
d_U	0.0453***	0.0508***	0.0444***	0.0453***	0.0508***	0.0444***	0.0453***
	(0.0141)	(0.00519)	(0.00727)	(0.00750)	(0.00469)	(0.00468)	(0.00432)
Last 15 I	0.0247***		0.0247***	0.0247***		0.0247***	0.0247***
	(0.00643)		(0.00581)	(0.00582)		(0.00768)	(0.00741)
Last 15 M	0.0162***		0.0162*	0.0162*		0.0162***	0.0162***
	(0.00614)		(0.00847)	(0.00849)		(0.00467)	(0.00450)
Last 15 U	0.0376***		0.0376***	0.0376***		0.0376***	0.0376***
	(0.00665)		(0.00847)	(0.00849)		(0.00958)	(0.00924)
Tokens Sent	0.000252			0.000252**			0.000252***
	(0.000212)			(0.000106)			(0.0000656)
Tokens Not Invested	0.0000434			0.0000434			0.0000434
	(0.000307)			(0.0000904)			(0.0000463)
N	1980	1980	1980	1980	1980	1980	1980

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficients of neither *Tokens Sent*, nor *Tokens Not Invested*, appear statistically significant. The sign and magnitude of all other estimated parameters remains roughly unchanged.

The following three columns in Table 5 report results from analogous regressions with and without *Tokens Sent* and *Tokens Not Invested*. In these regressions, standard errors are not clustered. The last three columns in Table 5 report results from the same analysis with standard errors clustered at the process level. Recall that our experimental design entails a draw of 15 Wiener processes, each utilized twice.[8] It is at this process level that we cluster in the last three columns.

Results are similar across all these specifications. One exception is the coefficient on our altruism proxy, *Tokens Sent*, which appears statistically significant, if very small, when we do not cluster standard errors or cluster at the process level. Nonetheless, about 60% of participants give 0 tokens,

[6] In the majority and unanimity treatments, this variable represents the group average tokens not invested.

[7] In the majority and unanimity treatments, this variable represents the group average tokens sent.

[8] In each session, the last 15 processes corresponded to a reflection of the first 15. Therefore, we effectively have two observations for each process in each of our treatments.

and more than 80% give less than 50 tokens. Given the estimated parameter value, this variable has limited ability to explain the variations in stopping posteriors we observe.

In most of our regression analysis, we cluster standard errors at the individual level. Table 6 presents analogous regression results with no clustering and process-level clustering. The first two columns consider process-level clustering at the group level. The next columns focus on individual stopping posteriors, as those discussed in the main text. The fixed-effects regression cannot be presented with process-level clustering as the panels are not nested within clusters.

Table 6: Decreasing Thresholds - Alternative Clustering

	Posterior					
	Process Level Clustering		No Clustering			
	OLS Regression		Ordinary Regression		Fixed Effects Regression	
	All Rounds	Last 15 Rounds	All Rounds	Last 15 Rounds	All Rounds	Last 15 Rounds
Constant	0.785***	0.806***	0.785***	0.806***	0.777***	0.821***
	(0.00929)	(0.00912)	(0.00463)	(0.00542)	(0.00426)	(0.00585)
d_M	-0.0303***	-0.0372***	-0.0303***	-0.0372***		
	(0.00973)	(0.0109)	(0.00819)	(0.00958)		
d_U	0.0347***	0.0431***	0.0347***	0.0431***		
	(0.00521)	(0.00677)	(0.00819)	(0.00958)		
Last 15 I	0.0247***		0.0247***		0.0299***	
	(0.00768)		(0.00546)		(0.00521)	
Last 15 M	0.0162***		0.0162**		0.0218***	
	(0.00467)		(0.00796)		(0.00775)	
Last 15 U	0.0376***		0.0376***		0.0431***	
	(0.00958)		(0.00796)		(0.00794)	
Slow I	-0.0648***	-0.0576***	-0.0648***	-0.0576***		
	(0.0171)	(0.0177)	(0.00547)	(0.00793)		
Slow M	-0.0774***	-0.0736***	-0.0774***	-0.0736***		
	(0.0160)	(0.0170)	(0.00798)	(0.0116)		
Slow U	-0.0440*	-0.0271	-0.0440***	-0.0271**		
	(0.0217)	(0.0227)	(0.00798)	(0.0116)		
Time I					-0.000651***	-0.00110***
					(0.000149)	(0.000180)
Time M					-0.00133***	-0.00167***
					(0.000397)	(0.000553)
Time U					-0.000517***	-0.000700***
					(0.000184)	(0.000240)
N	1980	990	1980	990	1980	990

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The only noticeable difference from the results presented in the Appendix of the main text is the weakening, or loss of statistical significance, of *Slow U* under process-level clustering.

4.9 Session Interactions

Since our group treatments entail a limited number of sessions, one may worry that interactions within sessions are driving our results. We now illustrate various ways by which our results appear robust to the session-partitioning in our data.

4.9.1 Probit with Session-Level Clustering

Our results remain statistically significant even when clustering standard errors at the session level, as shown in Table 7 here. The table presents estimates from analysis analogous to that presented in the main text. As can be seen, results are virtually identical.

Table 7: Probit Regression (Session-level Clustering)

	$P(Vote)$					
	Individual	Majority	Unanimity	Individual	Majority	Unanimity
Posterior	5.357***	5.149***	5.690***	5.071***	3.787***	5.463***
	(0.375)	(0.813)	(0.0938)	(0.178)	(0.605)	(0.0388)
Time	0.242***	0.798***	0.333***	0.313***	0.673**	0.328***
	(0.0840)	(0.134)	(0.111)	(0.0666)	(0.279)	(0.125)
Slope				0.137***	0.132***	0.0475***
				(0.0179)	(0.0223)	(0.00877)
Standard Dev				-0.142	0.626***	0.350***
				(0.317)	(0.173)	(0.0592)
Constant	-4.980***	-4.626***	-5.263***	-4.891***	-3.880***	-5.192***
	(0.251)	(0.394)	(0.0460)	(0.112)	(0.316)	(0.0862)
N	7865	6772	11113	6824	5301	9660

Standard errors in parentheses
Session-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We next show that there is a statistically significant difference between the estimated parameters across treatments. Table 8 reports regression results that speak explicitly to the difference between the estimated parameters for the majority and unanimity treatments, clustering at the session level. Instead of running the regressions separately, we define D_M as a dummy variable that equals 1 for observations from the majority treatment and 0 otherwise. We interact this dummy variable with all parameters of interest. Thus, for all intents and purposes, the regressions whose results are reported in Table 8 are nearly identical to those underlying Table 7 here. However, Table 8 allows direct conclusions on the statistical significance of differences between the treatments.

Table 8: Unanimity Compared to Majority (Session-level Clustering)

	$P(Vote)$	
Posterior	5.690***	5.463***
	(0.0861)	(0.0357)
Time	0.333***	0.328***
	(0.102)	(0.115)
Slope		0.0475***
		(0.00805)
Standard Dev		0.350***
		(0.0543)
Constant	-5.263***	-5.192***
	(0.0422)	(0.0792)
Posterior $\times D_M$	-0.540	-1.677***
	(0.776)	(0.575)
Time $\times D_M$	0.465***	0.345
	(0.163)	(0.289)
Slope $\times D_M$		0.0842***
		(0.0227)
Standard Dev $\times D_M$		0.276
		(0.173)
D_2	0.638*	1.313***
	(0.376)	(0.310)
N	17885	14961

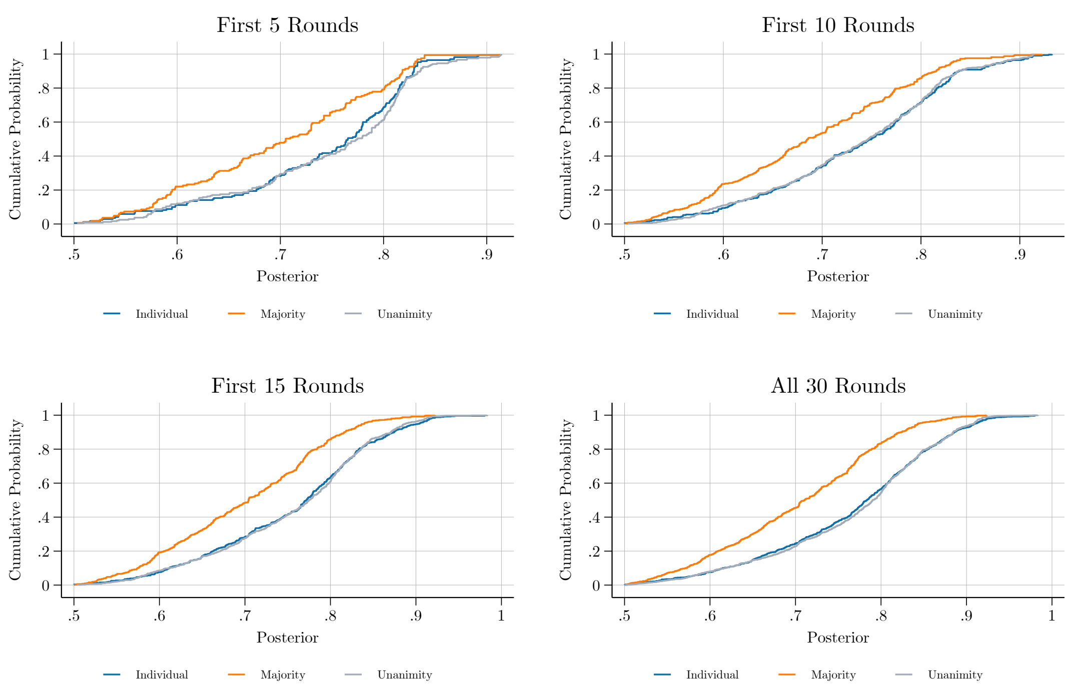
Standard errors in parentheses
Session-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As can be seen, when regressing $P(Vote)$ on posteriors and time, the majority treatment exhibits significantly more pronounced reactions to time than the unanimity treatment. When regressing $P(Vote)$ on the posterior, time, slope, and standard deviation of the process, we see that results from the majority treatment significantly differ from those under the unanimity treatment in the initial intercept, the reaction to the posterior, as well as the reaction to the slope of the process. In all cases, we have p-values lower than 0.01. Similar differences can be seen for the individual and majority treatments.

4.9.2 Presence of Effects in Initial Rounds

The effects we capture are present from the very start of our sessions, before individuals have had a chance to interact extensively with others in the session, as shown in Figure 7. The figure shows that the comparisons across treatments is present in the first 5, first 10, first 15, as well as in all 30 rounds.

Figure 7: Effect Present in Initial Rounds



Across all our treatments, the distributions of pivotal choices similarly exhibit striking stability over time.

Table 9: Probit Regression (Session-level Clustering)

	Posterior									
	Individual-level Clustering					Session-level Clustering				
	Round: 1	Round: 1-2	Round: 1-3	Round: 1-4	Round: 1-5	Round: 1	Round: 1-2	Round: 1-3	Round: 1-4	Round: 1-5
D_M	0.0319^{*}	0.0377^{***}	0.0440^{***}	0.0330^{***}	0.0440^{***}	0.0319	0.0377^{*}	0.0440^{**}	0.0330^{**}	0.0440^{***}
	(0.0177)	(0.0133)	(0.0122)	(0.00984)	(0.00953)	(0.0236)	(0.0208)	(0.0186)	(0.0112)	(0.0103)
Constant	0.627^{***}	0.683^{***}	0.696^{***}	0.715^{***}	0.700^{***}	0.627^{***}	0.683^{***}	0.696^{***}	0.715^{***}	0.700^{***}
	(0.0146)	(0.0113)	(0.0102)	(0.00859)	(0.00813)	(0.0205)	(0.0181)	(0.0162)	(0.00973)	(0.00837)
N	115	229	344	458	573	115	229	344	458	573

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9 establishes statistical significance of the comparisons depicted for early rounds in Figure 7. We utilize data from the first one, two, three, four, or five rounds. In the regressions underlying the table, D_M represents a dummy variable as before: it equals 1 for observations from the majority treatment. We report results for both individual-level and session-level clustering.

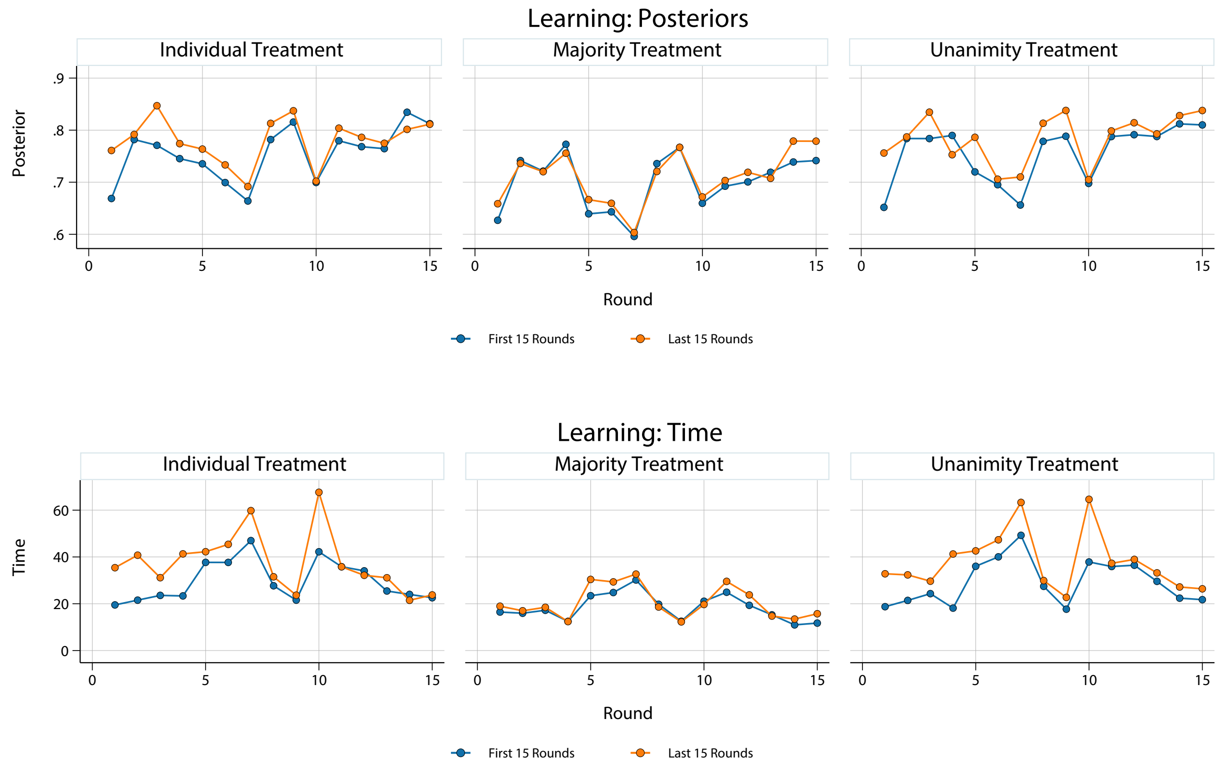
When clustering errors at the individual level, we reach a $p < 0.01$ significance level utilizing only the first two rounds of data. When clustering errors at the session level, we reach a $p < 0.05$ significance level utilizing data from the first three rounds, while a $p < 0.01$ significance level is reached by utilizing data from the first five rounds. Thus, the fundamental patterns in our data are robustly present from the very start of our sessions, when interactions with others in the session are severely limited.

4.9.3 Learning across Treatments

Next, we compare individual-level choices made in each dynamic treatment in the first and last 15 rounds. Recall that in the last 15 rounds, participants experienced the same sample paths (albeit mirrored). Thus, we have a highly controlled environment to study learning. Figure 8 depicts how choices evolve.

As can be seen, in terms of both the posteriors at which participants cast their votes and the time they took, there is a remarkable similarity between the individual and unanimity treatment. Furthermore, the sample path itself heavily influences both the choice of posterior and time. For example, in sample path 10, which is repeated 15 rounds later as sample path 25, we see that participants in the individual and unanimity treatment spend a lot of time. Consequently, due to the decreasing thresholds we identify, they submit their votes with lower posteriors.

Figure 8: Learning across Treatments



4.9.4 Group Influence

Participants do not appear to be influenced by prior group members' choices in our treatments, as shown in Table 10 here. The table reports results from an individual-level fixed effects regression of the posterior with which a participant cast a vote in round t, on the round number, $Round$, as well as the difference between the posterior with which they cast a vote in round $t-1$, from the mean posterior with which other group members cast a vote in round $t-1$, denoted $Difference_{t-1}$. The $Constant$ captures the average treatment fixed effect. Each column represents a separate fixed-effects regression, for the different dynamic treatments, for either all 30 rounds, the first 15 rounds, or the last 15 rounds (we use individual-level clustering, since the clustering level must match the fixed effects level).

If participants are influenced by their group members, we expect the estimated coefficient on $Difference_{t-1}$ to be negative and statistically significant. For example, when a participant voted with a posterior higher than the rest of the group, group influence would lead the difference to be positive: if the participant is influenced by the group's decisions, she should decrease the posterior

Table 10: Dynamic Group Effects

	Individual			Majority			Unanimity		
	All	First 15	Last 15	All	First 15	Last 15	All	First 15	Last 15
Round	0.00211***	0.00595***	0.00108	0.00136***	0.00240*	0.00414***	0.00201***	0.00380***	0.00364***
	(0.000461)	(0.000951)	(0.000956)	(0.000404)	(0.00132)	(0.000711)	(0.000301)	(0.000708)	(0.000776)
$Difference_{t-1}$				0.0198	0.0170	-0.0227	0.0433	0.0352	-0.0322
				(0.0437)	(0.0759)	(0.0500)	(0.0283)	(0.0334)	(0.0393)
Constant	0.734***	0.707***	0.755***	0.679***	0.678***	0.607***	0.742***	0.731***	0.700***
	(0.00714)	(0.00761)	(0.0220)	(0.00661)	(0.0114)	(0.0164)	(0.00482)	(0.00602)	(0.0179)
N	1020	510	510	728	339	389	1392	672	720

Standard errors in parentheses
Individual-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

with which she casts her vote in the next round. Table 10 here illustrates that there is no such adjustment.

4.10 Demand for Agency: Second and Third Voters

Table 11 presents a regression similar to the one discussed in the the main text. The dependent variable here is the difference between the posterior of the third and second vote. Since in the majority treatment only two votes are required for a decision to be made, this regression utilizes data only from the unanimity treatment and the simulated individual treatment.[9]

Table 11: Stopping Posteriors: Third and Second Voters

	$(p_3 - p_2)$
Constant	0.186***
	(0.0204)
d_U	-0.0794
	(0.0498)
p_2	-0.548***
	(0.0492)
$p_2 \times d_U$	0.100*
	(0.0561)
Last 15	0.0228***
	(0.00765)
Last 15$\times d_U$	-0.00942
	(0.00780)
Slow	-0.00672
	(0.0221)
Slow$\times d_U$	-0.00266
	(0.0126)
N	330518

Standard errors in parentheses
Process-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As can be seen, there is no statistically significant difference between the intercepts in the simulated individual treatment and in the unanimity treatment. In contrast, the coefficient of

[9] Since p_1 can take values between 0.5 and 1, before running the regression, we re-normalize all the values of p_1 by subtracting 0.5. Thus, the intercept corresponds to the additional posterior the third voter places when the second voter cast a vote with a posterior of 0.5.

$p_2 \times d_U$ is statistically significant at the 10% significance level. Thus, the unanimity treatment is associated with a slightly flatter slope than the simulated individual treatment. However, since its intercept is also lower, the difference between the two remains rather small.

4.11 Declining Best Responses

The logic is simple: a rational voter grouped with individuals exhibiting declining thresholds effectively faces a constrained problem. The induced constraint reduces the voter's continuation value, but does not affect the value of voting immediately and stopping. Hence, interacting with other participants with declining thresholds increases the relative appeal of voting earlier.

Using the notation introduced in Section 2, consider a rational voter whose individually optimal decision thresholds are (g^*, G^*). This voter interacts with two other group members. Assume, for the moment, that the first of these two follows a threshold strategy (with either stationary or declining thresholds for either alternative) and that she is always the first to vote. Suppose the third voter has declining decision thresholds. She votes for A at time t when $\theta \geq G(t)$ and votes for B at time t when $\theta \leq g(t)$, with $G(t) > g(t)$ for all t, $G(t)$ and $-g(t)$ decreasing in t, where $G(\hat{t}) = G^*$ and $g(\hat{t}) = g^*$ for some $\hat{t}$.

Under majority rule, after time $\hat{t}$, a rational voter is constrained by the presence of the third voter. In effect, she is constrained to choose A between $[0, G(t)]$ and B between $[g(t), 0]$. Let $(\tilde{g}(t), \tilde{G}(t))$ denote an optimal solution to this constrained stopping problem. Although the solution does not have a closed-form expression, it is straightforward to see that $\tilde{g}(t) \geq g^*$ for all t. Let $V(t, \theta)$ and $\widetilde{V}(t, \theta)$ denote the value functions of the unconstrained and constrained stopping problems, respectively, at time t with log-likelihood ratio θ. Recall that (g^*, G^*) is the unique optimal solution in the unconstrained case, and that any feasible stopping strategy in the constrained case is also feasible in the unconstrained case. If $\tilde{g}(t) < g^*$ at some t, then for $\theta \in (\widetilde{g}(t), g^*)$, we must have $\widetilde{V}(t, \theta) < V(t, \theta) = D(\theta)$, where $D(\theta)$ stands for the immediate payoff from voting for B. This inequality contradicts the supposition that $(\tilde{g}(t), \tilde{G}(t))$ is optimal, as the voter can choose to stop at t when the belief is θ. Similar reasoning shows that $\tilde{G}(t) \leq G^*$ for all t.

This discussion suggests that a pivotal voter who expects others to use declining thresholds may choose lower standards of accuracy, leading to faster decisions under majority rule.[10] By contrast,

[10] A similar logic applies even if the rational voter is initially the first voter with the lowest demanded accuracy.

under unanimity rule, the rational voter is pivotal and unconstrained by the decisions of the other two voters after $\hat{t}$. Hence, she will simply follow her individual optimal rules after $\hat{t}$. In this case, the group decision will be decided by the third-order statistic of the individual thresholds and coincide with that of a simulated group that comprises the same three voters.

References

Chan, J., Lizzeri, A., Suen, W., and Yariv, L. (2018). Deliberating collective decisions. *Review of Economic Studies*, 85(2):929–963.

Dvoretzky, A., Kiefer, J., Wolfowitz, J., et al. (1953). Sequential decision problems for processes with continuous time parameter. testing hypotheses. *The Annals of Mathematical Statistics*, 24(2):254–264.

Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, 112(2):631–645.

Wald, A. (1945). Sequential method of sampling for deciding between two courses of action. *Journal of the American Statistical Association*, 40(231):277–306.

If other voters' thresholds eventually lead to this voter becoming pivotal, then the first voter will optimally vote to stop earlier than her individually optimal threshold.