Individual and Collective Information Acquisition: An Experimental Study*

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

October 10, 2022

Abstract

Many committees—juries, political task forces, etc.—spend time gathering costly information before reaching a decision. We report results from lab experiments focused on such information-collection processes. We consider decisions governed by individuals and groups and compare how voting rules affect outcomes. We also contrast *static* information collection, as in classical hypothesis testing, with *dynamic* collection, as in sequential hypothesis testing. Several insights emerge. Static information collection is excessive, and sequential information collection is non-stationary, producing declining decision accuracies over time. Furthermore, groups using majority rule yield especially hasty and inaccurate decisions. Nonetheless, sequential information collection is welfare enhancing relative to static collection, particularly when unanimous rules are used.

*We thank Marina Agranov, Roland Benabou, Daniel Benjamin, Tim Cason, Stephen Morris, Salvatore Nunnari, and Wolfgang Pesendorfer for very helpful discussions and feedback. We gratefully acknowledge the support of NSF grants SES-1629613 and SES-1949381.

†Department of Economics, Duke University pellumb.reshidi@duke.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 Introduction

1.1 Overview

Information acquisition precedes a variety of important decisions. The pioneering work of Wald (1947) illustrated that collecting costly information dynamically, using *sequential hypothesis testing*, is more efficient than deciding statically on the sample size, using *classical hypothesis testing*.[1] In many settings, information is indeed collected sequentially—agencies like the FDA and the EPA design their evaluations sequentially, with each stage contingent on the one preceding it: juries recall testimonies through dynamic deliberations; job-market candidates often proceed to advanced interview stages contingent on their success in earlier ones. Nonetheless, despite the efficiency of sequential sampling, institutional or time constraints often restrict decisions to occur statically—up or down tenure decisions; conclusions from field experiments set in advance due to logistical constraints; or rapid military responses. In addition, many important information-collection decisions are made by groups: expert committees, juries, department faculty, or military intelligence and combat officers. Do groups behave differently from individuals when collecting information? Does the aggregation procedure within groups matter? In practice, are there large benefits from the utilization of dynamic rather than static information-collection protocols? There is a dearth of data addressing these questions. Our aim is to fill this gap.

This paper reports results from an array of lab experiments inspecting both static and sequential information collection by individuals and groups. Furthermore, we consider groups operating under different aggregation procedures. In our design, individuals and groups trade off information accuracy and costs, either statically or dynamically. Since participants face the same incentives as individuals or as groups, optimal choices should, theoretically, be independent of the aggregation rule. In particular, in our sequential settings, individuals and groups should use a constant cutoff posterior, as in Wald (1947), to determine when information collection comes to a halt. Similarly, in static settings, individuals and groups should choose the same sample size regardless of the aggregation rule.

We find the following patterns in our data. First, on average, sequential information collec-

[1] Wald (1947)'s insight was perceived so profound that, in the introduction to his book, it is noted that "Because of the usefulness of the sequential probability ratio test in development work on military and naval equipment, it was classified Restricted within the meaning of the Espionage Act."

tion leads to decision accuracies that come close to those that are predicted by the theory, but these accuracies decline over deliberation time. This pattern suggests that individuals display time-declining thresholds instead of the stationary ones predicted by the theory. Second, static information collection is excessive. Third, groups exhibit markedly different behaviors than individuals, with majority rule leading to faster (and less accurate) decisions. Despite these deviations from the theoretically-optimal behavior, we find that welfare is higher when information is collected dynamically, especially under unanimity.

Our results have implications for institutional design when information collection is an important component of decision making. Using committees, rather than individuals, can be beneficial even if increasing the size of the decision body does not affect the overall information available. Nevertheless, the rules governing how collective decisions are made and the information protocol in place—static or dynamic—need to be customized in tandem.

At the core of our experimental design is the following decision problem. There are two ex-ante equally likely states, A or B. States can represent a guilty or an innocent defendant, a safe or unsafe drug, an investment that is profitable or not, etc. Ultimately, each participant must guess the state of the world and gets rewarded when correct. Each state is associated with a Brownian motion. The drift is μ when the state is A and $-\mu$ when the state is B. The Brownian motion's variance is state-independent. As time goes by, the realized sample path of the Brownian motion becomes increasingly informative about the underlying state. There is a flow cost of information collection. Whenever information collection terminates, participants know the posterior probability that the state is A and submit their guess. Naturally, the optimal guess corresponds to the more likely state. Our focus is on the trade-off pertaining to information collection: waiting longer before making a decision increases accuracy, but comes at a cost.

We consider both static and dynamic information-collection procedures. The dynamic setting implements the sequential sampling setup.[2] Participants track the evolution of the Brownian path and can stop at any time to submit their guess. The static setting is designed to resemble the dynamic setting while emulating the classical hypothesis testing setup. Participants determine, at the outset, the time horizon during which they collect information by observing the Brownian path. They then see the path unravel for their desired time, get informed of the ultimate posterior over

[2] Specifically, our setup mimics that of Dvoretzky et al. (1953).

states, and make their guess.[3]

In our benchmark treatments, decisions are made by individuals, as in the canonical paradigms. In additional treatments, decisions are made in groups. When in a group, we consider two commonly-used institutions: majority and unanimity. In the dynamic setting, group members decide whether to stop or continue information collection at each point in time. Under majority, whenever two members agree on a guess, information collection terminates for the group, and the majority guess is submitted. Analogously, under unanimity, whenever all members agree on the guess, information collection terminates, and that guess is implemented. In the static setting, group members all submit their desired information-collection horizon at the outset. Under majority rule, the median time is implemented for the group, whereas under unanimity, the maximal time is implemented. In all our group treatments, information is public: group members are privy to the same information. Furthermore, group members receive the same payoff, derived from the common costs accrued during the group's information-collection period and the group's guess accuracy.

Our individual treatments offer a natural benchmark for the basic predictions emerging from the classical statistical information-collection procedures. In the dynamic setting, it is optimal to stop whenever there is sufficient confidence in guessing the state, namely when the posterior belief exceeds a time-independent threshold—at that point, the costs of additional information exceed its benefits. With our parameters, the optimal threshold posterior is 0.81. In our experimental treatments, individuals' mean posterior at decision time is relatively close to that predicted by theory, standing at 0.77. However, participants do not use time-independent thresholds: we see decreasing thresholds over time, with participants becoming more lenient as time passes.[4] In particular, contrary to theoretical predictions, quicker decisions tend to be more accurate. As we discuss in our literature review, this observation is consistent with a wide neuroscience literature documenting a similar pattern using perception tasks. Our data are unique since, by design, we directly observe the posterior probabilities participants see over time. In the static setting, our parameters are such that the optimal information-collection horizon is 30 seconds. In our experiments, on aver-

[3]Ex-ante, it is not clear which decision protocol is simpler. Static information collection asks for only one choice at the outset, whereas dynamic information collection requires continuous monitoring of the posterior's evolution. Nonetheless, static information collection provides more limited feedback than dynamic information collection on the relation between sample size and decision accuracy.

[4]Although we see participants making decisions at lower posteriors as time goes by within rounds, this pattern does not extend across rounds: participants' mean stopping posteriors are higher in the second half of our sessions.

age, individuals choose 42 seconds, a choice that is 40% higher than is optimal. Since individuals under-collect information in the dynamic setting, the wedge in decision accuracy between dynamic and static information collection is not as large as theoretically predicted. Nevertheless, when accounting for information costs, dynamic information collection yields higher expected payoffs, as Wald (1947) suggested.

To identify group effects, disjoint from preference-aggregation effects, our groups are, by design, homogeneous.[5] Theoretically, efficient group outcomes should coincide with individual outcomes. Specifically, threshold posteriors and waiting times should be independent of the decision rule. Nonetheless, we find that participants in groups behave differently from participants making decisions in isolation. This behavior depends on both the information-collection protocol and the voting rule in place.

Do group interactions always affect behavior? Not necessarily. In dynamic treatments under unanimity, participants' behavior is virtually identical to that seen in our individual treatment.[6] Still, this group neutrality does not hold under majority, as participants make much quicker decisions than in our individual treatment. It is challenging to find a force rooted in group dynamics that explains this hastiness under majority while having no effect in unanimity. We suggest individuals may display a *demand for agency*: a desire to have a say in the vote. In dynamic treatments, under unanimity, all group members get to express their preferences before a decision is made—a demand for agency should have no effect. In contrast, under majority, a decision can be taken before all members had cast their vote and a demand for agency introduces a potential race. Another possible explanation relates to our finding of declining decision thresholds. Although the theoretically optimal stationary thresholds are individually optimal, and the best response to any symmetric stationary threshold, best responding to decreasing thresholds by other group members involves faster decisions.

Since participants' behavior differs from the theoretical benchmark, in principle, so could the welfare rankings between static and dynamic treatments. Nonetheless, we find that sequential information collection always outperforms static information collection in terms of expected welfare.

[5] Chan et al. (2018) analyze theoretically dynamic group information collection in the presence of heterogeneity. We view the homogeneous setting as a natural empirical starting point.

[6] The group outcomes differ since we see heterogeneity in individual behaviors and unanimous decisions reflect the preferences of the most demanding group member.

In fact, the highest welfare is observed in the case of groups that collect information sequentially and use unanimity rule. The impact of the information-collection protocol is nuanced, however, and not always as pronounced at theory predicts. In particular, decisions made by groups using majority are more accurate when information is collected statically rather than dynamically. This is a consequence of the excessive information collection in our static treatments combined with the hastiness observed in our dynamic majority treatments. Thus, when democratic decision bodies affect a large population, so that information costs are inconsequential to society, static information collection may be beneficial.

1.2 Related Literature

The problem of testing statistical hypotheses is an old one. Its origin can be traced back to Thomas Bayes, who provided the well-known formulation of posterior probabilities of event "causes" in the 18th century. Classical hypothesis testing has been used, formally or informally, for centuries, see Stephan (1948). It came of age with the development of statistical hypothesis tests by Neyman and Pearson (1933), who showed that the likelihood ratio test is the most powerful hypothesis test for a given data set. Examples abound for its uses. See, for example, Greene (2018).

Sequential sampling, proposed by Wald (1945, 1947), introduced the idea of collecting data dynamically. With each piece of data, a likelihood ratio test is performed to determine whether more observations are needed to accomplish a desired level of statistical confidence. When data come at a cost, Wald's method offers efficiency gains over its static counterpart—when data is collected in increments, a researcher can condition additional data collection on what had already been observed. Sequential sampling has been used widely to describe how individuals collect information, more on that below, and to guide researchers in the creation of databases, see Dominitz and Manski (2017) and references therein.

Recent theoretical work has investigated how groups approach the deliberative process, linking information acquisition with ultimate decisions. Persico (2004), Martinelli (2006), and Gerardi and Yariv (2007, 2008) investigate environments in which information collection by a committee is "static," reminiscent of classical hypothesis testing. In those models, each individual can acquire a costly signal about a payoff-relevant state. The aggregation process then introduces free-riding motives. This contrasts with our setting, where any information collected by the group is public, with its costs equally shared.

Strulovici (2010), Chan et al. (2018), and Henry and Ottaviani (2019) consider environments in which information collection is sequential: the committee decides at each date whether to continue acquiring costly information, or stop and choose an alternative. In particular, Chan et al. (2018), which our dynamic group treatments mimic, as well as Henry and Ottaviani (2019) and McClellan (2021), build on the literature on sequential hypothesis testing that started with Wald (1947).

In terms of experimental work, an old literature in psychology considers settings similar to ours when looking at individual choices. Becker (1958) studies dynamic information collection of 8 participants and suggests non-stationary decision heuristics. Edwards and Slovic (1965) also study dynamic information collection, where 10 participants gain either no information or very precise information at each step, so that updating is trivial. They report evidence for under- and over-collection. Green et al. (1964) report data from 36 participants who make static information-collection choices. They, too, report both under- and over-buying of information. Perhaps the most related to our work is the paper of Fried and Peterson (1969), who document behavior of 20 participants choosing information statically and dynamically. Like us, they see non-stationary behavior in the dynamic setting and show that, despite deviations from optimal behavior, dynamic information collection yields higher expected values. While this literature serves as a natural backdrop to our study, there are several important differences. First and foremost, we investigate how *groups* using different rules collect information statically and dynamically. Methodologically, beyond the substantially larger sample size and more detailed statistics, our design allows us to disentangle Bayesian updating from optimization heuristics by providing participants the posterior at every point in the dynamic treatment. Furthermore, while these previous studies commonly present participants with a finite number of informative signals they can acquire, the volume of information participants can collect in our setting is effectively unbounded, thereby making the problem easier to interpret through the lens of the classical models of information collection.

Within economics, there is a large literature that studies how individuals collect and process information statically. Many papers consider the collection of information when agents have non-instrumental motives, for example seeking confirmatory information as in Fischer et al. (2005) or ego-promoting information as in Eil and Rao (2011). Relatively few papers study experimentally how individuals trade off precision of payoff-relevant information and its costs, which is at the heart of the classic hypothesis testing paradigm. Ambuehl and Li (2018) elicit valuations of payoff-relevant information structures. They show that valuation of information under-reacts to increased informativeness, but that individuals value information that may yield certainty particularly highly. Hoffman (2016) uses a field experiment in which business experts are compensated for their guess of the price and quality of actual websites. Participants can acquire a costly signal before deciding. He also finds that participants underpay for strong signals and overpay for weak signals. Our static treatments add to this literature by illustrating how both individuals and groups resolve the accuracy-cost trade-off.[7]

Several papers inspect individual dynamic search behavior experimentally, see Gabaix et al. (2006), Brown et al. (2011), Caplin et al. (2011), and references therein. In these experiments, participants also spend resources over time in the hopes of identifying a good alternative, but the underlying optimization problem is quite different from ours. Chen and Heese (2021)'s experiment resembles our individual dynamic treatment. However, their focus is on the ethical valence of the alternatives.[8]

The neuroscience literature has produced a rich body of work that inspects binary perceptual tasks. Response times are often interpreted as costly, turning the problem into a sequential sampling one, often termed the drift-diffusion model. Much of the focus of this literature concerns the association between correct choice rates and response times, see for instance Swensson (1972), Luce et al. (1986), Ratcliff and Smith (2004), and Ratcliff and McKoon (2008). The main finding emerging from this literature is that quick decisions tend to be more accurate. This insight is in line with our observation of declining thresholds in the dynamic treatments: as time passes, our participants stop information collection with less certainty on the correct choice. An important contrast with these studies is that we observe—in fact, provide—the posterior probability that any choice is correct over time. This allows us to speak directly to new theories of dynamic choice that have emerged recently, see Baldassi et al. (2020) and Fudenberg et al. (2018).

[7] Several studies inspect information collection in strategic settings different from ours. Elbittar et al. (2020) and Bhattacharya et al. (2017) consider information aggregation settings in which individuals acquire private information, Szkup and Trevino (2021) explore information collection in the context of global games, while Gretschko and Rajko (2015) focus on auctions.

[8] Canen (2017) provides some field evidence on voters sequentially collecting information prior to elections. Interestingly, the idea of using sequential experimental *designs* has been suggested in various contexts, see El-Gamal and Palfrey (1996), Chapman et al. (2019), Imai and Camerer (2018), and references therein.

2 Experimental Design

A description of the interface and sample instructions are available in the Online Appendix. At the core of our experimental design is the choice of the amount of information to acquire prior to making a binary decision. There are two possible states: A and B. Although labeled neutrally in the lab, these can stand for a guilty or innocent defendant in the jury context, a good or bad policy in the political context, a profitable or unprofitable investment in a finance context, etc. At the start of each period, one of the states is chosen at random with probability $1/2$. Participants ultimately need to guess the state and are paid according to the correctness of their guess. In the lab, participants receive \$2 for a correct guess and nothing otherwise.

Before making their guess, participants have access to information that evolves according to a continuous-time Wiener process. The process has state-independent instantaneous variance σ^2, but state-dependent drift. When the state is A, the drift is μ; When the state is B, the drift is $-\mu$. To produce reasonable expected round durations, throughout our treatments, $\mu = 0.84$ and $\sigma^2 = 1$. Naturally, our experimental software provides an approximation of the continuous setup, where the interface is updated five times a second.[9]

Our treatments vary across two dimensions: whether information acquisition decisions are static or sequential and whether choices are made by individuals, groups using majority rule, or groups using unanimity rule. Each participant took part in only one of the resulting six treatments.

In what follows, we start by describing our sequential treatments. The design of these treatments guided our design of the static treatments, which are described next.

Sequential Sampling Our dynamic treatments mimic the sequential-sampling environment of Dvoretzky et al. (1953). In these treatments, participants observe information evolve over time and, at each instant, can guess A, B, or wait for further information by choosing W. Information comes at a flow cost of 40 cents a minute.

In the treatment in which individuals make decisions on their own—the *individual dynamic* treatment—a round ends as soon as a participant makes an A or B guess.

In our group treatments, participants are randomly matched to form groups of three in each

[9] We use a single parameterization, which allows for a rich study of information collection in a variety of settings with a manageable number of treatments. This trade-off between richness of parameters and environments is present across experimental topics: repeated games, auctions, voting, etc., see Kagel and Roth (1995). Our design is amenable to different parameter choices, offering a natural future step.

round. Information is public: all individuals in the group observe the same information. A round ends as soon as a quorum of q individuals agrees on an A or B guess. In the *majority dynamic* treatment, $q = 2$. In the *unanimity dynamic* treatment, $q = 3$. As long as a quorum has not been reached, participants can change their decisions between A, B, and W at any time. Throughout, participants can see the choices of other group members.

Static Sampling Our static treatments mimic the setting of the *classical hypothesis testing* environment. At the beginning of each round, participants decide on the amount of time they want to spend collecting information. As in the dynamic treatments, information costs are fixed at 40 cents a minute.

When individuals make decisions independently—the *individual static* treatment—they observe the information evolve for the amount of time that they chose.[10] Their guess is then automated to reflect the state that is more likely given the information collected: either A or B.[11]

Our static-sampling group treatments are analogous to those corresponding to the dynamic treatments. In each round, participants are randomly re-matched into groups of 3. At the outset of each round, participants submit simultaneously their desired waiting time. In the *majority static* treatment, the resulting group waiting time is the median desired waiting time of group members: this is the minimal time at which a majority of group members would agree to stop information collection. It is the maximal desired waiting time of group members in the *unanimity static* treatment: this is the minimal time at which all group members would agree to stop information collection. As in the individual treatment, participants observe the information evolve for the amount of time chosen by the group. The group guess, A or B, is again automated.

Feedback and Payments In all treatments, the feedback at the end of each round contains participants' payoffs and other group members' choices when relevant. In groups, all members are paid the same amount, incorporating the accuracy of the group's guess and information costs.

Each treatment was preceded by two practice rounds, followed by 30 payoff-relevant rounds. Participants were ultimately paid for 20 randomly-selected rounds out of these 30.

[10] This design was chosen for two reasons. First, we wanted to maximize comparability with the sequential-sampling treatments. Second, we wanted to offer participants sufficient learning opportunities.

[11] The guess is automated in order to reduce noise in our data. Because participants' guesses in the individual dynamic treatment best respond to the information 98% of the time, it is unlikely this restriction impacts our qualitative results.

Information Processes The 30 information processes participants experienced in the experimental rounds were identical across treatments. To select these processes, we randomly generated 15 sample paths with the parameters specified above. These processes are "representative" in that the mean, median, and five quintiles of the theoretically-optimal sequential stopping times match those of the underlying distribution (see the following section for a description of the theoretical predictions). These processes correspond to the first 15 real rounds in each treatment. The last 15 processes in each treatment were derived by generating the reflected "mirror images" of the first 15 processes. Namely, whenever the realized state in the original process is A (or B), it is B (or A) in the reflected process. Furthermore, at any time t, if the original process indicates a probability p that the state is A, the reflected process indicates a probability $1 - p$ that the state is A. The reflected processes were used in the same order as the original processes. In that way, participants effectively faced the same 15 decision problems twice during a session. This design element allows us to evaluate learning in a highly controlled fashion.[12]

The evolution of a Wiener process provides continuous information on the likelihood of either state. Nonetheless, the Bayesian calculus necessary to deduce this likelihood is non-trivial. The difficulty this calculus introduces is orthogonal to our investigation.[13] To mitigate the impacts of participants' limitations in statistical analysis, our design directly displays the evolution of the *probability* that the state is A (or B).

Auxiliary Elicitations At the end of each session, participants completed two risk-elicitation tasks as in Gneezy and Potters (1997). Namely, participants were provided with 200 tokens that they had to allocate between a safe investment, returning token for token, and a risky investment with a mean higher than 1 and a non-trivial variance (e.g., one paying 2.5 the amount invested with probability 50%). In addition, participants took part in two dictator-games, one in which the amount of tokens transferred was translated $1:1$ and one in which the amount of tokens transferred was doubled for the recipient. Participants were paid for one randomly-chosen risk-elicitation task and one randomly-chosen dictator game.[14]

[12] As we soon describe, the evolution of the process was depicted through a uni-dimensional scale capturing posterior probabilities updated over time. Identifying repetitions is extremely unlikely: it would require the memorization of many ordered values and the realization that they are mirrored.

[13] It is well known that lab participants are frequently challenged by statistical updating, see for instance the survey of Benjamin (2019).

[14] We elicited duplicate responses to allow for measurement-error correction as suggested in Gillen et al. (2019).

Table 1: Participants and Rounds

	Dynamic		Static	
	Participants	Rounds	Participants	Rounds
Individual	34	1,020	31	930
Majority	48	480	48	480
Unanimity	48	480	45	450

Summary The experiments were run at the Princeton Experimental Laboratory for the Social Sciences (PExL) with 254 participants. We conducted at least four sessions for each group treatment, with at least 12 participants in each. Table 1 summarizes our treatments and the corresponding volume of participants.[15] The experimental software was programmed using oTree (Chen et al., 2016).

Approach to Data Analysis As may be expected, participants' behavior changes during early experimental rounds as they learn about the problem. We see no evidence for substantial learning in later rounds. For details, see the Appendix. Throughout the paper, we present figures aggregated across all experimental rounds as those displayed appear virtually identical when we use either the full data or the last half of our sessions. Regression results are presented for data corresponding to all rounds in the text, and for the last 15 rounds in the Online Appendix. The qualitative messages remain the same.[16] We also discuss individual- and session-level heterogeneity in the Appendix. Risk attitudes and altruism proclivities do not appear to play an important role in explaining patterns in our data, even after measurement-error correction. We therefore do not include data from these elicitations in our main specifications. See the Online Appendix for related analyses, which also allow for various levels of clustering.

3 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

[15] Given our grouping protocol, the number of rounds per participant in our group treatments is three times lower than in the individual treatment.

[16] Recall that the processes participants encounter in the first and second half of each session are equivalent.

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, $\mu' = 1.4$.

3.1 Static Treatments

In order to obtain the optimal wait time in the static setting, we need to compute the probability of guessing the true state correctly for any chosen time t. This probability can be shown to be given by the following expression[17]

$$\int_0^\infty \frac{1}{\sqrt{4\pi\mu' t}} e^{-\frac{(x-\mu' t)^2}{4\mu' t}} dx = \frac{1}{2}\left(\operatorname{erf}\left(\frac{\sqrt{\mu' t}}{2}\right) + 1\right).$$

In the static setting, a risk-neutral agent maximizes

$$\max_t \frac{1}{2}\left(\operatorname{erf}\left(\frac{\sqrt{\mu' t}}{2}\right) + 1\right) - c\,t.$$

The optimal wait time is then

$$t^* = \frac{2\,W\left(\frac{(\mu')^2}{32\pi c^2}\right)}{\mu'},$$

where $W(\cdot)$ is the Lambert W function (i.e., $W(x) = w$ if and only if $x = we^w$).

With our parameter values $t^* = 0.49$. Since one unit of time in the lab is one minute, this optimal wait time translates to 29.58 seconds.[18] A consequence of our parameter choices is that the optimal duration is not overly extreme, implying that errors are unlikely to be one-sided.

Consider now a group of $n > 1$ homogeneous agents who choose their desired search times simultaneously. The group then collects information for a duration corresponding to either the median or the maximal specified time. As before, the group guess corresponds to the more likely

[17] If $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal, the error function $\operatorname{erf}(\cdot)$ is defined so that $\operatorname{erf}(x) = 2\Phi(\sqrt{2}x) - 1$.

[18] A discussion of this setting in the presence of risk aversion is presented in the Online Appendix.

state given the posterior that is generated by the collected information. Group members are (identically) rewarded as in the one-agent setting.

The utilitarian efficient equilibrium for the group corresponds to the optimal search time described above, namely 29.58 seconds. Furthermore, this choice is weakly dominant for any agent, regardless of the strategies other agents in the group utilize.

3.2 Sequential Treatments

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[19]

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

The corresponding first-order condition with respect to the lower boundary is then[20]

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

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, in the dynamic version, a risk-neutral agent should wait

[19]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.

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

Table 2: Aggregate Behavior

	Dynamic Treatment				Static Treatment			
	Mean Posterior		Mean Time Waited		Mean Posterior		Mean Time Waited	
	All Rounds	Last 15	All Rounds	Last 15	All Rounds	Last 15	All Rounds	Last 15
Individual	0.77	0.78	33.56	37.55	0.75	0.75	41.69	40.45
	(0.003)	(0.005)	(0.687)	(1.12)	(0.004)	(0.006)	(0.561)	(0.824)
Majority	0.73	0.73	23.07	24.38	0.74	0.74	36.25	34.48
	(0.002)	(0.003)	(0.335)	(0.51)	(0.003)	(0.005)	(0.326)	(0.515)
Unanimity	0.82	0.84	46.71	53.68	0.76	0.75	40.46	37.77
	(0.002)	(0.003)	(0.724)	(1.11)	(0.004)	(0.005)	(0.343)	(0.547)
Theory	0.81		39.03		0.72		29.58	

Standard errors in parentheses

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.

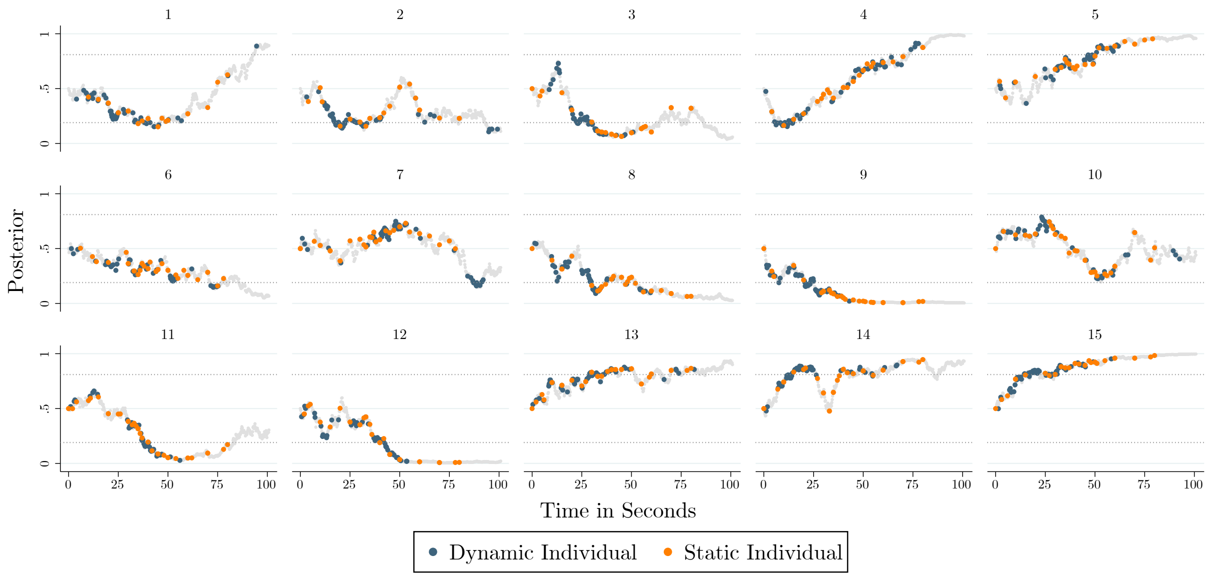
The utilitarian efficient equilibrium for the group, under both majority and unanimity, corresponds to the optimal search policy described above, namely utilizing a threshold of 81%. In our experiment, participants make group decisions with other individuals who do not choose the exact 81% threshold. 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.

4 Broad Patterns of Behavior

Table 2 displays an aggregate overview of some of our results. It displays the estimated mean of the posteriors with which the pivotal vote has been cast, and the estimated mean time taken to cast the pivotal vote. As can be seen, our individual and majority dynamic treatments lead to less accurate decisions than theoretically predicted, whereas the unanimity dynamic treatment yields outcomes that are extremely close to those theory predicts.[21] Furthermore, the majority dynamic

[21] The path of posteriors intersects the stopping threshold from below. If participants implemented a threshold with trembles drawn independently over time, our estimated threshold would underestimate the "true" threshold: a downward tremble in the threshold is more likely to lead to termination than an upward tremble. In the Appendix, we allow for such noise and show that the resulting bias is very small in our data.

Figure 1: Pulling the Trigger: Individual Treatments



treatment corresponds to the least amount of waiting, an observation we shall return to.

Differences between threshold posteriors in the data and those predicted by theory may, at first blush, appear small. Nonetheless, these differences translate to fairly large differences in wait times. For instance, the unanimity dynamic treatment leads to double the wait time in majority dynamic treatment. This is a direct consequence of the convexity of precision costs— the marginal time required to attain a given increase in precision is increasing in the level of the precision.

Static treatments yield excessive waiting relative to that predicted by theory. Again, the majority-rule treatment generates the hastiest decisions, though differences are not significant.

Contrary to theory, mean decision times are longer in the static treatments than in the dynamic treatments for both individuals and groups using majority. Moreover, the differences between mean posteriors at decision times in static and dynamic treatments are not as large as theory predicts.

Figure 1 depicts the evolution of posteriors and the choices made in each of our 15 processes in the individual treatments, both static and dynamic. Our use of identical processes across treatments allows for such a direct comparison. In order to simplify the presentation, each panel aggregates observations from two reflected processes (for example, panel 1 corresponds to the first and sixteenth process, panel 2 to the second and seventeenth process, etc.). The Figure illustrates the point at which individuals “pulled the trigger” and the pivotal vote was cast.

The Figure suggests some important themes that appear in our more detailed analysis below. First, it is apparent that decisions are heterogeneous, corresponding to a spectrum of stopping posteriors. Second, many observations are close to optimal. In fact, many participants stop information collection at the theoretically-predicted posterior accuracy (corresponding to the horizontal dashed lines within each panel). In the dynamic setting, participants clearly respond to information in that decisions are more clustered around higher posteriors. Third, individuals in the dynamic treatment become more lenient, requiring less accuracy to stop, the longer they wait: they display decreasing thresholds. Consider, for example, process 10. Several individuals decide late in the process, when posteriors are close to 50%, despite choosing not to stop at earlier points, when posteriors were close to 80%. Last, because in the static treatment individuals cannot condition their choices on the history, the resulting decision posteriors are far more dispersed.[22] For instance, in processes 9 and 12, some static choices take place at extreme posteriors (close to zero) that had already stabilized for some time. Earlier stopping would have been preferable if agents had been able to condition their behavior on the history. In contrast, in processes 2 (around 50 seconds) and 14 (around 35 seconds), some static decisions terminate at posteriors of around 50% in regions where no dynamic decisions terminate.

The analogous figure for our majority and unanimity treatments appears in the Appendix. Results are similar: we see more leniency over time in the dynamic treatments, and more decisions at extreme posteriors—either low or high—in the static treatments.

In what follows, we analyze the behavior that underlies these initial observations. The next section describes behavior in our dynamic treatments. The section that follows offers a comparison with their static counterparts.

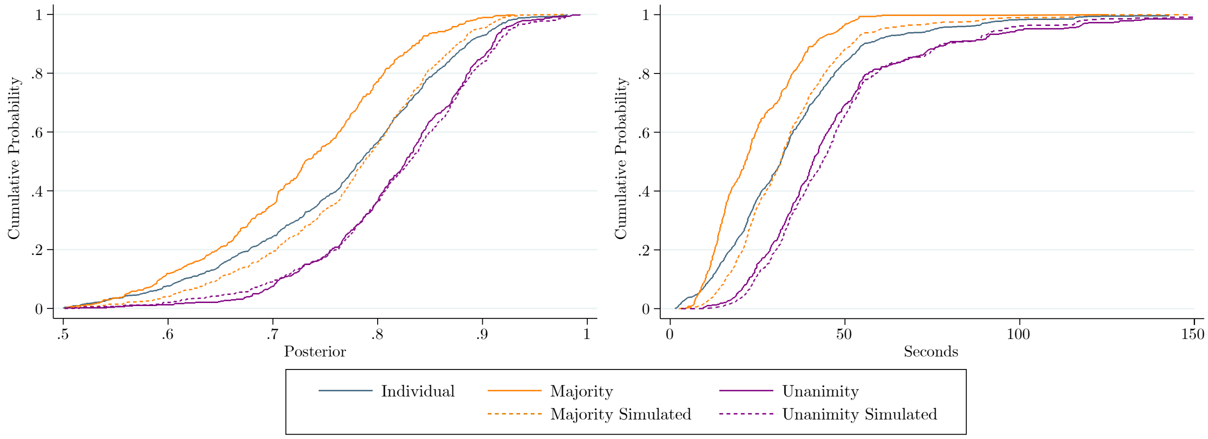
5 Sequential Information Collection

5.1 The Impacts of Decision Procedures

For each of our dynamic treatments, Figure 2 displays the cumulative distribution functions of the posteriors when the decision was made (the left panel) as well as the associated decision times (the right panel). We see a substantial impact of the governing decision rule. Distributions can be

[22]We return to a discussion of posterior dispersion in Section 7.

Figure 2: Dynamic Treatment CDFs



ordered via first order stochastic dominance, with the unanimity dynamic treatment yielding the highest-accuracy decisions and taking the longest to conclude, and the majority dynamic treatment yielding the least-accurate and hastiest decisions. In particular, the averages presented in Table 2 are not principally driven by outliers. As discussed in Section 3, theory suggests we should not observe any differences in outcomes among voting rules. However, theory also predicts a single threshold posterior, whereas we observe substantial heterogeneity in behavior in the individual treatment. Given this heterogeneity, it is natural to ask whether the differences we observe across our dynamic treatments are simply a mechanical consequence of the grouping of three random individuals that respond heterogeneously to the task at hand. Specifically, groups governed by majority rule decide according to the second order statistic, whereas groups governed by unanimity rule decide according to the third order statistic.

To assess whether the differences we see among these treatments are purely mechanical, we simulate hypothetical groups of three participants by drawing data from our individual treatment.[23] Figure 2 presents the resulting cumulative distribution functions from these simulated groups, alongside the distributions we observe in our data. The additional accuracy granted by groups using unanimity appears to be a purely mechanical phenomenon.[24] In contrast, groups using majority rule yield substantially less accurate and hastier decisions than those of simulated groups

[23] For each round, we randomly group the 34 participants in our individual treatment into 11 groups of 3 participants, randomly discarding one. We do so $1,000$ times. Across all 30 rounds, $330,000$ groups are therefore simulated.

[24] A two-sided Kolmogorov-Smirnov test fails to reject the hypothesis that these distributions, the simulated and observed unanimity group decisions, are identical. One possible concern is that observations generating these figures are correlated. This raises questions about the validity of standard statistical tests for comparing these distributions; see additional analysis in the Online Appendix. We soon use regression analysis, with adequate error clustering, to statistically determine what affects decisions.

using majority, suggesting that hasty majority choices are not the pure consequence of a mechanical aggregation effect.[25] We will discuss this phenomenon in more detail in Section 5.5 below.

5.2 Non-stationary Behavior

We now assess the determinants of when participants decide to terminate information collection. Theoretically, the probability of voting should be 1 when the posterior reaches its theoretical threshold value of 0.81, and 0 for any lower posterior. In particular, the probability of voting should respond only to observed posterior probabilities, not to the time that has passed, to features of the sample path, or to choices of other group members.

We describe the behavior of our participants by presenting results from a probit regression. The left-hand side variable captures whether a participant has voted, and the main explanatory variable is the posterior. In the left panel, as an additional explanatory variable we include the time (in minutes) to allow for time dependence in voting outcomes. To allow for the possibility of path dependence, in the right panel, we also include features of the sample paths. Specifically, we divide each round into (non-overlapping) 5-second time intervals. Within each 5-second window, we record our left-hand side variable—whether a vote was cast; our explanatory variables—the posterior and time at the end of the window; as well as the slope and standard deviation of the sample path.[26] We utilize data up until participants cast their individual votes. Thus, an observation consists of an indicator of whether the participant cast their vote by the end of the 5-second interval, the posterior at the end of the interval, the time at the end of the interval, as well as the slope and standard deviation of the process during that same 5-second interval. Table 3 reports the corresponding coefficient estimates.

[25]A two-sided Kolmogorov-Smirnov test rejects the hypothesis that these distributions are identical, though the caveat regarding such tests still holds. We use regression analysis below to make a stronger statistical case for the difference between group treatments.

[26]The slope corresponds to the average posterior gain per minute calculated using the 5-second window. To make the ranges of posteriors and standard deviations comparable across 5-second windows, we normalize the standard deviation through a multiplication by 5. Different time windows, of 3,...,7 seconds, yield similar results. So do regressions focusing on the last 15 rounds of each session. See the Online Appendix for details.

Table 3: Probit Regression

	$P(Vote)$					
	Individual	Majority	Unanimity	Individual	Majority	Unanimity
Posterior	5.357***	5.149***	5.690***	5.071***	3.787***	5.463***
	(0.400)	(0.426)	(0.406)	(0.402)	(0.478)	(0.427)
Time	0.242**	0.798***	0.333***	0.313***	0.673***	0.328***
	(0.120)	(0.179)	(0.111)	(0.109)	(0.189)	(0.110)
Slope				0.137***	0.132***	0.0475*
				(0.0360)	(0.0338)	(0.0253)
Standard Dev				-0.142	0.626**	0.350*
				(0.212)	(0.275)	(0.203)
Constant	-4.980***	-4.626***	-5.263***	-4.891***	-3.880***	-5.192***
	(0.344)	(0.291)	(0.312)	(0.332)	(0.367)	(0.335)
N	7865	6772	11113	6824	5301	9660

Standard errors in parentheses

Individual-level clustering

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

The left panel of Table 3 indicates that decisions to cast a vote are responsive to posteriors, with higher posteriors naturally leading to increased voting probabilities. In addition, *stopping decisions are not stationary*: controlling for posteriors, the more time passes, the more likely agents are to make a decision. For example, in the individual treatment, waiting for one additional minute is equivalent to an approximate increase of 4.5 percentage points in the observed posteriors ($0.045 \times 5.357 \sim 0.242$): the probability an individual casts a vote with a hypothetical posterior of 0.80 at the outset approximately coincides with the probability an individual casts a vote with a posterior of 0.76 after one minute has passed. Furthermore, time appears to have a stronger impact on the likelihood of making a decision when groups use majority rule.

In the right panel, we also include features of the sample paths. The coefficients corresponding to posteriors and time passed change only slightly. In addition, the coefficient corresponding to *Slope* is positive and significant. This implies that after a brief period in which posteriors increase rapidly, a decision is more likely, particularly for individuals and groups using majority rule. The coefficient corresponding to *Standard Dev* is barely significant, however. That is, recent variation in posteriors has a limited effect.

For an alternative approach in which we analyze observed stopping posteriors directly, see the Appendix. Similar conclusions emerge.

Our non-stationarity results are related to the results of Brown et al. (2011). They experimentally study a stationary job-search problem with a known distribution of wage offers. Reported reservation wages decrease over time. They consider two potential explanations for this phenomenon: non-stationary time discounting, and a "sunk-cost fallacy" whereby agents set reservation wages in response to cumulative costs. Both effects are present in their treatments, although the first is more pronounced. As mentioned above, our participants also react to features of the process itself, a phenomenon that could not be examined in Brown et al. (2011). Furthermore, although within a round later decisions are made at lower posteriors, this is not the case between rounds. In fact, the average stopping posteriors, as well as durations, are higher in sessions' second half.[27] In addition, one of our findings is incompatible with agents placing exaggerated weight on cumulative costs: as we documented in Section 4, agents wait excessively in our static treatments.

Our finding that the probability of casting a vote is increasing in time, even while controlling for the posterior, is also connected to the drift-diffusion model (DDM)—see, e.g., Swensson (1972), Luce et al. (1986), Ratcliff and Smith (2004), and Ratcliff and McKoon (2008). As mentioned above, this literature finds that quick decisions tend to be more accurate. An important contrast with these studies is that we observe—in fact, provide—the posterior probability that any choice is correct at each point in time. This allows us to speak directly to new theories of dynamic choice that have emerged recently. The explanation provided by Fudenberg et al. (2018) for the relationship between speed and accuracy relies on decision-makers' uncertainty about payoffs, which translates into uncertainty about the process and leads to optimal non-stationary behavior. In our setting, the problem is inherently stationary, and the only uncertainty is about which of two drifts governs the process. In our data, experience does not significantly reduce the degree to which thresholds are decreasing, suggesting that it is unlikely that "subjective uncertainty" about the process is what drives behavior. Furthermore, as mentioned, stopping behavior responds to the sample path itself, behavior that cannot be explained with a pre-determined (potentially time-variant) threshold, as in Fudenberg et al. (2018).[28]

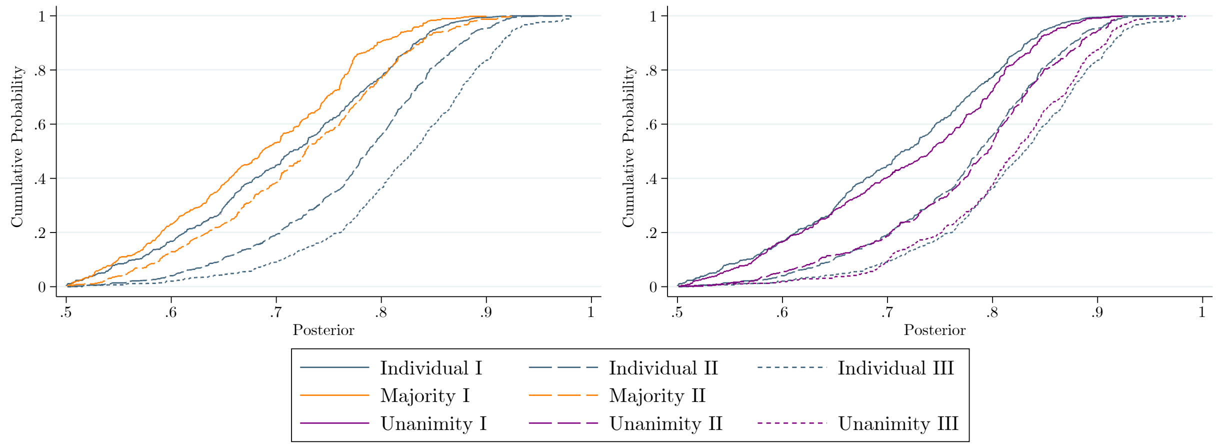
5.3 Voting First, Second, and Third

We now discuss how the patterns of behavior compare between our treatments. In Figure 3 we present the distributions of posteriors corresponding to the first and second votes in the majority treatment (in the left panel), and the distributions of posteriors corresponding to the first, second,

[27] Recall that the second 15 rounds utilize the same processes as the first 15 rounds, only mirrored. Thus, the increase in the observed stopping posteriors cannot be an artifact of features of the processes themselves.

[28] Baldassi et al. (2020) offer an axiomatic approach to the drift-diffusion model. McClellan (2021) derives non-stationary threshold posteriors as the consequence of agency frictions. Strack and Viefers (2021) report on path dependence in a related search setting. See also references therein.

Figure 3: Dynamic Treatment CDFs by Vote Order



and third votes in the unanimity treatment (in the right panel). Alongside these distributions, we present analogous distributions for simulated groups of three generated from the individual treatment via the procedure described in Section 5.1.

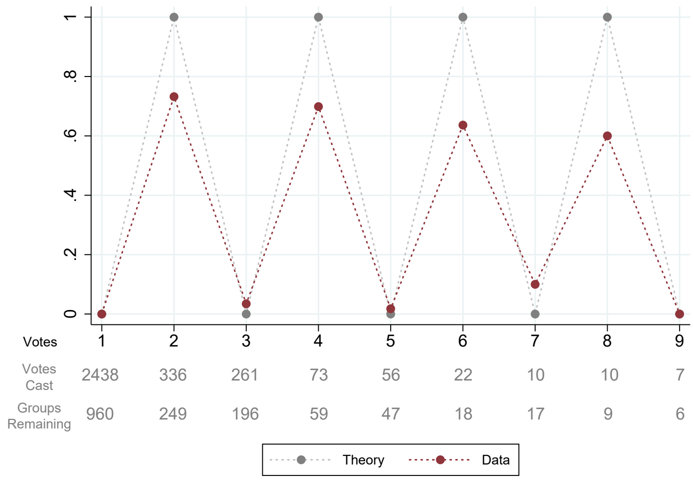
An implication of our discussion in Section 5.1 is that the third-order statistic from the individual simulated treatment is very close to the distribution of the third and pivotal voter from the unanimity treatment. The right panel of Figure 3 confirms this finding and reveals that this similarity also holds for the first and second voter. Therefore, this figure reinforces the idea that individual voter behavior under unanimity is very similar to behavior of individuals deciding in isolation, and that the differences in outcomes under unanimity are exclusively due to the aggregation rule acting on heterogeneous individuals.

For the majority treatment, the left panel of Figure 3 demonstrates that hasty behavior is not only a characteristic of the second (and pivotal) voter; the first voter appears to be hasty as well. Both the first- and second-order statistics from the simulated individual treatment stochastically dominate the observed distributions corresponding to the first and second voters from the majority treatment. Interestingly, the distribution of second voters under majority is very similar to the distribution of first voters in the individual simulated treatment, a point we soon return to.

5.4 Individual-Level Dynamic Choices

In a group setting, as long as a pivotal vote has not been cast, participants can change their vote from supporting one alternative, say A, to either waiting (W) or supporting the other alternative

Figure 4: Share of W Votes in Dynamic Majority and Unanimity Treatments

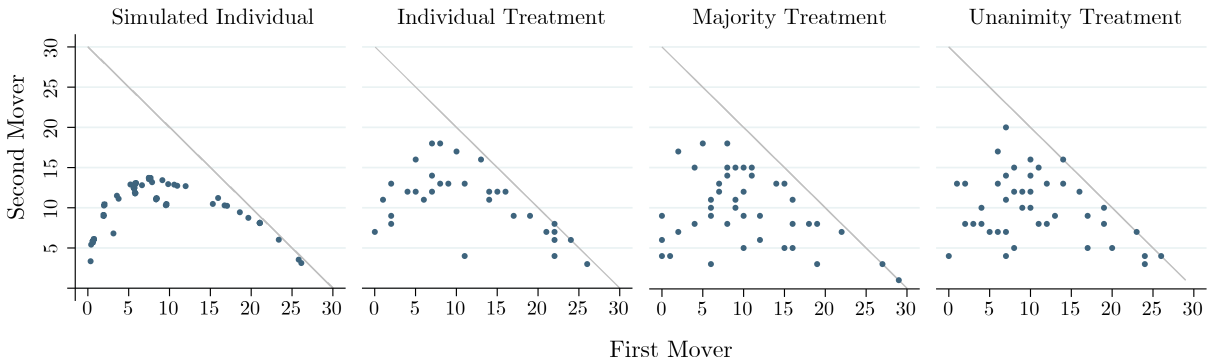


B (following the notation of Section 2). In both our group treatments, roughly 85% of rounds end with each participant casting at most one vote. Therefore, the first votes in a group yield a good approximation of order statistics for the group treatments. They also offer the theoretically valid way to compare group and individual treatments—in the individual treatment, the first vote terminates information collection. Nonetheless, if participants use symmetric cutoff strategies, a transition from a choice of A to a choice of B, or vice versa, should entail a range of posteriors in which W is chosen. When the posterior that the state is A (or B) is high enough, A (or B) is chosen; when there is no strong support for either state, W is chosen.

Individual behavior resembles these theoretically-predicted patterns, if imperfectly, as shown in Figure 4. The figure illustrates the individual vote sequence in our majority and unanimity dynamic treatments. The vertical axis represents the fraction of W choices. Theoretically, the first vote is for A or B. With a cutoff strategy, any switch away from the initial vote is to a W choice, followed by an A or B choice, and so on. The figure depicts this benchmark. Behavior in the data is close to the theoretical benchmark, albeit with one difference: participants do not utilize the W option between A and B choices to the full extent that the theoretical benchmark indicates.

We now discuss the degree to which there are persistent “types” in our sample. As Figure 5 shows, we do not see substantial persistence in vote orders of individuals: there are very few participants who are always first, always second, or always third to vote. Each dot in a figure’s panel corresponds to one participant. The figure displays the number of rounds participants ended up

Figure 5: Individual-Level Dynamic Voting Order Across Rounds



as first movers, second movers, or third movers. The left-most panel, titled Simulated Individual, displays the outcome of 330,000 simulations of groups of three participants from our individual treatments, formed at random, and the resulting order of votes—using the observed times at which these participants voted—see details in Footnote 23. The second panel, corresponding to our individual treatment, displays a similar figure using only one random grouping of participants. The two right panels report the order of votes in our majority and unanimity treatments.

Some individuals appear to be willing to vote with little information, and often move first. Some demand a high level of certainty and often vote last. However, we see few individuals who always move first, second, or third.[29]

5.5 Explaining Hasty Majority Decisions

Why are decisions under majority so hasty while unanimity decisions are not? There are several known behavioral channels the literature has suggested for explaining behavior changes in groups that do not appear consistent with our data. For example, the experimental literature has explored many ways by which individuals display other-regarding preferences in groups; see Cooper and Kagel (2016). However, it is challenging to reconcile individuals' hasty behavior under majority with their seemingly unaltered behavior under unanimity. Another classic group effect pertains to individuals' desire for *diffusion of responsibility*; see Darley and Latané (1968). A desire to diffuse responsibility corresponds to individuals actively preferring not to affect the group's choice.

[29] In the Appendix, we also show the distribution of mean decision posteriors by individuals. While in our individual and unanimity treatments we see a substantial fraction consistently approximating the optimal behavior, there are no clear "types" that emerge otherwise.

Such a desire would push the second and third voters in the unanimity treatment to vote faster in order to avoid being pivotal. However, as shown in Figure 3, behavior by second voters under unanimity does not significantly differ from behavior by second voters in simulated groups based on our individual treatments. In what follows, we propose two complementary explanations for hasty majority decisions.

5.5.1 Demand for Agency

Recent work suggests individuals desire to influence outcomes. See, for instance, Fehr et al. (2013), Bartling et al. (2014), and Pikulina and Tergiman (2020). Such a *demand for agency* generates hasty majority decisions, but has limited effect on individuals acting on their on or in groups using unanimity. Indeed, when operating alone, or in a group using unanimity, agency is guaranteed—in both cases, a decision can only be made after each participant has cast a vote. In contrast, under majority rule, the group decision is made by two out of three group members, those who are first to vote. Thus, agency eludes a participant who pursues a more demanding threshold. There is, then, a non-trivial trade-off between desired accuracy and hastiness for the sake of agency.[30]

To evaluate the plausibility of a demand for agency, we start by inspecting remaining voters' responses to the first vote being cast. Our interest is in examining how behavior compares across treatments. Under majority rule, a demand for agency would introduce a race between the remaining two group members and thereby reduce the posterior at which the second vote is cast. Table 4 displays the results of a regression in which, within each combination of treatment, group, and round, we calculate the difference between the posterior at which the second vote was cast, and the posterior at which the first vote was cast.

The variables d_M and d_U are dummy variables corresponding to the majority and unanimity treatments, respectively. The variable p_1 stands for the posterior associated with the first vote cast in the group. Since p_1 can take values between 0.5 and 1, we re-normalize the values of p_1 by subtracting 0.5. Thus, the intercept corresponds to the *additional* accuracy required by the second voter when the first voter casts a vote with a posterior of 0.5. The variables $p_1 \times d_M$ and $p_1 \times d_U$ correspond to the interactions between p_1 and the corresponding treatment dummies,

[30] Another interpretation of the desire to influence is a *demand for pivotality*, whereby individuals wish to be decisive. Such a desire would lead second voters in the unanimity treatment to delay their votes in order to be more likely to be the pivotal voter. As already mentioned, this behavior is inconsistent with our data.

Table 4: Difference in Posterior: Second vs First Voter

	$(p_2 - p_1)$
Constant	0.193***
	(0.0176)
d_M	-0.154***
	(0.0338)
d_U	-0.00695
	(0.0205)
p_1	-0.607***
	(0.0639)
$p_1 \times d_M$	0.164***
	(0.0548)
$p_1 \times d_U$	0.00585
	(0.0288)
N	330960

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

allowing for different slopes across treatments.[31] To calculate the difference between the posteriors with which votes are cast, we rely on choices across different individuals. Thus, we cluster errors at the process level. We once more rely on simulating the first, second, and third votes from the individual treatment based on the procedure described in Section 5.

As can be seen, d_M and $p_1 \times d_M$ are both statistically significant at the 1% level, indicating a different slope and intercept for the majority treatment: in that treatment, the second voter places a lower "premium" on top of the posterior with which the first vote is cast. In other words, second voters are hastier under majority than they are under unanimity, or in the simulated groups based on the individual treatment. In contrast, there is no statistically significant difference between either the intercepts or the slopes of the unanimity and (simulated) individual treatments.[32]

The results of Table 4 are consistent with a demand for agency exhibited by later voters under majority rule. Furthermore, given the observed responsiveness of second voters to first-voters' choices under majority, there is a strategic reason for first voters to expedite their choices as well. Relatively impatient agents, who are likely to be first voters, can manipulate the pivotal threshold posteriors to be more in line with their preferences. If agents did not exhibit a demand for agency, a lenient group member, associated with a low threshold posterior, would have to

[31] Results remain virtually identical when controlling for learning and the time it takes to reach the theoretical stopping threshold in each sample path.

[32] In the Online Appendix, we compare the difference between the posteriors of the third and second vote in the unanimity treatment with that of the simulated individual treatment. There appears to be no statistically significant difference between the intercepts, whereas the slope of the unanimity treatment appears different only at the 10% significance level.

accept the higher, median threshold posterior utilized in the group. Instead, with a demand for agency, by expediting her choice, the more lenient member induces a hastier second vote—ideally, she would tailor the posterior at which she votes so that the pivotal vote would occur at precisely her desired threshold posterior. This is consistent with our observations: Figure 3 indicates that the distribution of posteriors when the second, pivotal votes are cast under majority closely approximates the distribution of posteriors when the first, most lenient votes occur under both the unanimity and the (simulated) individual treatments.

5.5.2 Best Responses to Declining Thresholds

At every instant, participants decide whether to act immediately or pay a cost to improve decisions in the future. If future decisions are constrained to be sub-optimal, a rational decision maker may best respond by voting quicker than otherwise. In fact, Chan et al. (2018) show that, when voters exhibit sufficient heterogeneity in preferences, under majority, some voters may rush to a decision in equilibrium: a current pivotal voter may foresee losing her status as pivotal in the future.

In our experiment, participants receive homogeneous payoffs, implying that individually optimal behavior involves constant and homogeneous thresholds. Of course, as we document in Section 5.3, participants' actual behavior is heterogeneous. This heterogeneity need not make a difference in understanding optimal behavior: as discussed in Section 3, as long as agents display symmetric thresholds, the best response is the individually optimal symmetric threshold regardless of the degree of heterogeneity. However, Section 5.2 shows that many participants also display declining thresholds. As we now illustrate, a rational participant who suspects her group members have declining thresholds may choose to be hastier under majority rule.

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 3, 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 $\hat{t}$, our 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. A 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.[33] By contrast, 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.

6 Static Information Collection

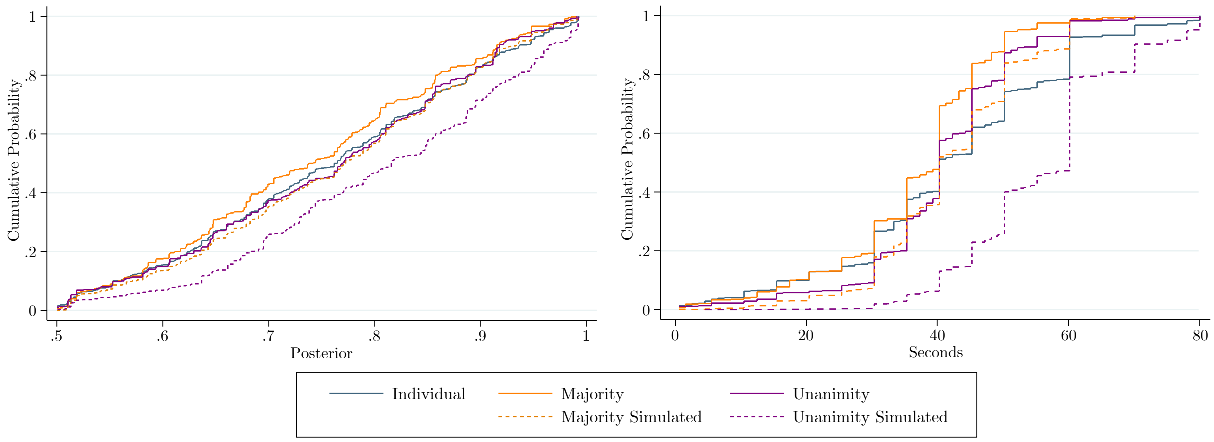
6.1 Group Level Distributions

In the dynamic treatments, we focused on the posterior probabilities associated with votes. In the static treatments, participants choose the duration of information acquisition at the outset. Thus, our focus shifts to the time chosen for information collection.

Figure 6 presents the cumulative distribution functions of chosen times across our static treat-

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

Figure 6: Static Treatment CDFs



ments, as well as the corresponding realized posterior probabilities.[34] In contrast to observations from the dynamic treatments, the distributions of chosen times across our static treatments appear similar to one another, although chosen times in the majority treatment are roughly first order stochastically dominated by those in both the individual and unanimity treatments: similar to our dynamic settings, groups using majority rule are hastier. These observations are in line with the average times chosen across treatments reported in Table 2.[35] The chosen-time distribution corresponding to the unanimity treatment second order stochastically dominates that of the individual treatment. That is, times chosen by individuals are more dispersed than times chosen by groups using unanimity rule.

The similarity between the distributions should be interpreted with caution. As in the dynamic case, the heterogeneity in individual choices implies that, were group members mimicking their choices as individuals, there would be differences across treatments; chosen times in our majority and unanimity treatments would correspond to the median and maximal preferred times within the groups. Therefore, to understand behavior in the group treatments, we simulate the distributions of choices in groups following the procedure described in Section 5. Namely, for each round, we form random groups of three individuals from our individual treatment and consider the median (majority simulated) and maximum (unanimity simulated) times within that group.

Under unanimity, it is the "most patient" group member who governs a group's decisions. It

[34] Participants predominantly specified an integer number of seconds, with some attraction to round numbers, corresponding to the apparent jumps in the distributions.

[35] In the Online Appendix, we use regression analysis with various levels of clustering to evaluate the differences across treatments.

is then unsurprising that the distribution of resulting simulated wait times under unanimity differs substantially from that corresponding to individual decisions. It also differs from our observed unanimity treatment, implying a non-mechanical difference between individual treatments and groups using unanimity rule. There is a similar non-trivial effect on groups using majority rule that is not mechanical: the simulated distribution does not coincide with those generated by observed group behavior.[36] Regression analysis in the Online Appendix confirms that the effects of both majority and unanimity rule are not purely a mechanical artifact. Furthermore, while we see some learning leading participants to select shorter times in the second half of our sessions, this learning is limited in scope and duration; see the Appendix for further details. In particular, throughout our experiments, both types of group treatments lead to *hastier* decisions than those generated by a purely mechanical aggregation effect.

The distribution of induced stopping posteriors are similar across our three treatments, although simulated groups using unanimity first-order stochastically dominate all other distributions. Why do we see differences in chosen times, but little differences in induced posteriors? In our static treatments, participants choose excessively long information-collection durations. Since expected induced posteriors are concave in these durations, the differences in chosen times between the treatments translate into smaller differences in the corresponding posteriors.

6.2 Individual-Level Static Choices

Figure 7 presents the distribution of the shortest (denoted by I, in analogy to Figure 3), median (denoted by II), and longest (denoted by III) chosen times for the (simulated) individual, majority, and unanimity static treatments.[37]

[36] The two-sided Kolmogorov-Smirnov test rejects the hypothesis that the distributions associated with the simulated and observed unanimity decisions, as well as the simulated and observed majority decisions, are identical.

[37] As for the dynamic treatments, when considering our individual treatment, we group participants into random groups of three and consider the shortest, median, and longest chosen times.

Figure 7: Static Treatment CDFs by Vote Order

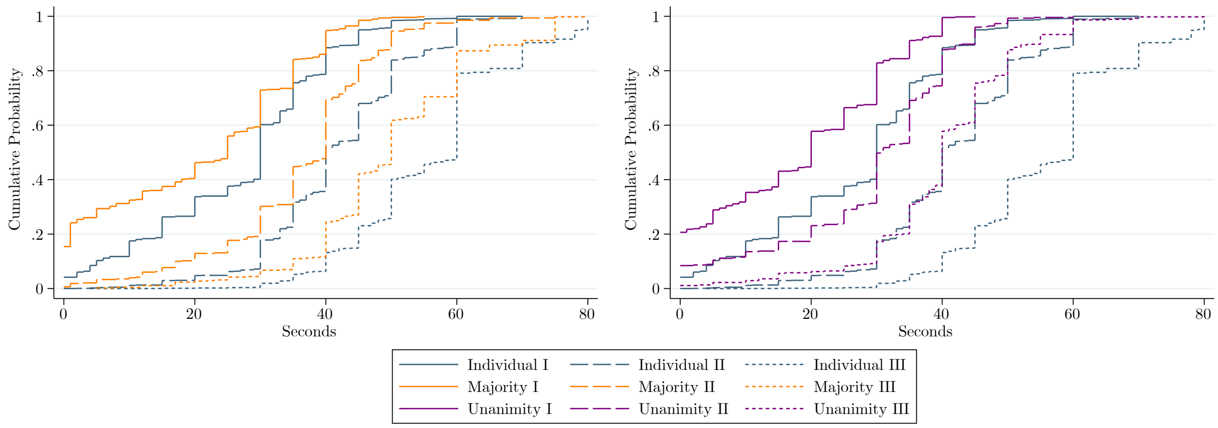


Figure 7 reveals a clear first order stochastic dominance relationship for the shortest, median, and longest times chosen across our treatments. The distributions of all three times corresponding to the unanimity treatment are dominated by those corresponding to the majority treatment, which are dominated by those corresponding to the simulated groups based on the individual treatment. In particular, behavior under both majority and unanimity differs from that in simulated groups. This echoes our conclusion that group effects are present and go beyond the pure mechanical effects driven by heterogeneity in our sample. In contrast to our dynamic treatments, all individual group members' votes are hastier under unanimity.

Figure 7 also suggests different degrees of heterogeneity among group members across our treatments. The times chosen within groups are closest in our unanimity treatment and furthest apart in our individual treatment. For instance, the distance between the median shortest and median longest chosen times under unanimity is approximately 20 seconds; the median distance is 30 seconds in simulated groups based on our individual treatment. Thus, more inclusive rules appear to generate a pressure for conformity.

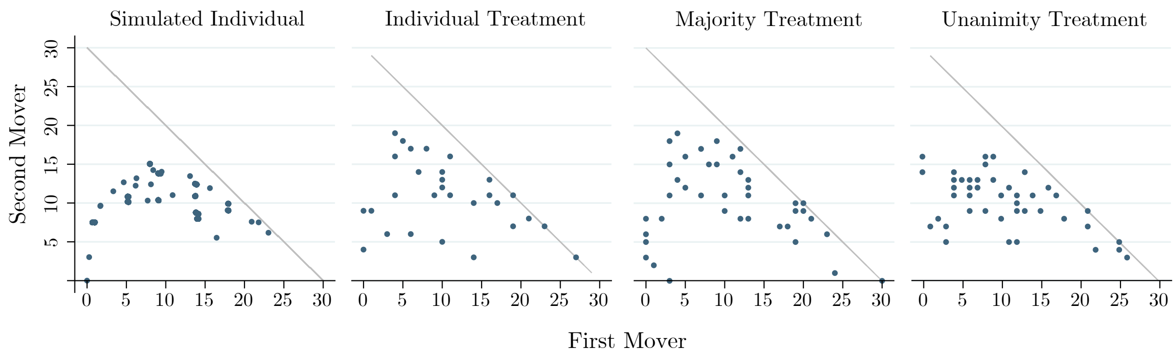
Figure 8 is the analogue of Figure 5.[38] As in our dynamic treatments, the figure suggests no substantial persistence in terms of "types," with very few individuals nearly always stating the lowest, the median, or the longest desired duration in their group.[39]

Certainly, there are several features that differ across our dynamic and static treatments. Most

[38] As for our dynamic treatments, we use 330,000 simulated random groups of three participants based on our individual treatment. We record each individual's order in the simulated group for each round, using their desired wait time in the data. The panel corresponding to the individual treatment displays similar results from one simulated grouping. The right-most panels correspond to the majority and unanimity treatment data.

[39] In the Appendix, we also show the distribution of mean desired wait times by individuals. There are no consistent "types" across our treatments.

Figure 8: Individual-Level Static Voting Order Across Rounds



notably, in dynamic treatments, participants observe processes evolve as they make their decisions, and can monitor other group members' choices over time. Instead, all decisions are made ex-ante in our static treatments. Both information-collection protocols lead to hasty majority decisions. However, in the static treatments, there is excessive information collection, even after many rounds of experience, and *both* majority and unanimity rules hasten participants' decisions significantly. In what follows, we assess the welfare implications of these differences.

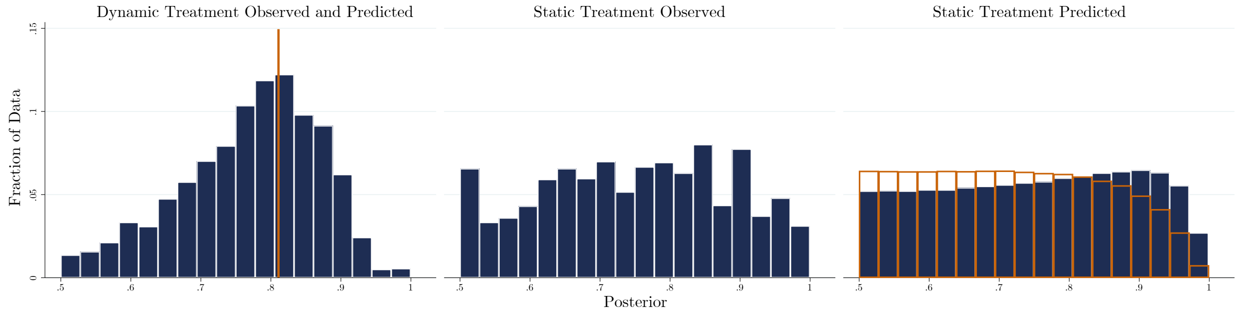
7 Performance

In this section, we compare the performance of individuals and groups to shed light on the impact of procedures and decision rules on ultimate outcomes, accounting for both decision quality and information costs.

Decision Accuracy In many settings, decisions made by a small group of individuals affect a large population—political decisions, jury verdicts, determinations of agencies such as the FDA or EPA, and so on. In such environments, a natural welfare criterion pertains to the accuracy of decisions: information-collection costs are born by only a minuscule fraction of the population. Theoretically, individuals and groups make the same choices regardless of the voting rule, and the only distinction is between static and dynamic information collection. The predicted accuracy is 0.81 in the dynamic setting, higher than the predicted accuracy of 0.72 in the static setting.

As Table 2 indicates, in our data, the accuracy differences between static and dynamic information collection are far smaller than theoretically predicted. Nonetheless, groups using unanimity

Figure 9: Posterior Histograms



and collecting information sequentially yield the most accurate decisions.

In contrast with theoretical predictions, under majority rule, static information collection yields more accurate decisions than its dynamic counterpart. This is a consequence of participants in our static treatments overshooting the theoretical benchmark. Although excessively costly for decision-makers, this overshooting generates greater accuracy. At the same time, in the dynamic setting, majority rule leads to particularly hasty decisions, which are therefore less accurate than theoretically predicted. There is a caveat to this comparison, however. By their nature, static decisions cannot directly target the posterior at which information collection stops. Thus, stopping posteriors tend to be more dispersed than when targeting an explicit accuracy threshold.

Figure 9 displays the distributions of posteriors in our dynamic and static treatments.[40] The left and middle panels display realized posteriors under the dynamic and static treatments, respectively (the vertical line on the left panel denotes the theoretically predicted stopping posterior of 0.81). The solid histogram on the right panel displays expected posteriors for the static treatments, conditional on their average observed waiting times (40 seconds). The hollow histogram on the right panel displays the theoretically expected posteriors for the static treatments with optimal wait time (29.6 seconds). As can be seen, the posteriors generated in our dynamic treatments (left panel) are far less dispersed than those of our static treatments (middle panel). This is in line with theoretical predictions. Thus, under majority, although static information collection may outperform the theoretical benchmark in terms of expected decision accuracy, it does run the risk of decisions being made with inconsistent precision levels.

[40]Separate figures for each of our decisions rules yield qualitatively identical patterns.

Table 5: Performance

	$\lambda^{benchmark}$		$\lambda^{expected}$	
	Performance Level	Relative Performance	Performance Level	Relative Performance
Individual Dynamic	0.655	0.863	0.651	0.841
	(0.00344)	(0.0191)	(0.00292)	(0.0162)
Majority Dynamic	0.650	0.832	0.648	0.820
	(0.00498)	(0.0277)	(0.00273)	(0.0152)
Unanimity Dynamic	0.662	0.902	0.660	0.889
	(0.00560)	(0.0311)	(0.00216)	(0.0120)
Individual Static	0.615	0.936	0.609	0.883
	(0.00325)	(0.0265)	(0.00213)	(0.0173)
Majority Static	0.617	0.952	0.615	0.935
	(0.00513)	(0.0417)	(0.00115)	(0.00933)
Unanimity Static	0.620	0.978	0.614	0.924
	(0.00444)	(0.0361)	(0.00123)	(0.00999)
N	3840	3840	3840	3840

Standard errors in parentheses
Individual-level clustering

Decision Accuracy and Costs We now turn to the evaluation of welfare accounting for costs; that is, welfare from the point of view of the committee making the decision. Indeed, in many environments—firms' hiring processes, publication review protocols, managerial investment choices—information costs are internalized by those affected by the ultimate decision.

We normalize the payoff for a correct guess to 1, the cost to 0.2, and divide the time waited in seconds by 60. Utilizing the posterior and time of the pivotal vote, we calculate the following performance measure, corresponding to average welfare

$$\lambda_{i,g}^{benchmark} = p_{i,g} - 0.2 \cdot t_{i,g}$$

where i represents a treatment, and g represents a particular group in a particular round within the treatment. We report the estimated mean of these performance measures under $\lambda^{benchmark}$ and *Performance Level* in Table 5 (first column).

In line with theoretical predictions, dynamic treatments generate higher performance than static treatments. These differences are statistically significant at any conventional significance level. Furthermore, there is a non-trivial interaction between the decision rule and information-collection protocols, albeit not statistically pronounced. In our dynamic treatments, the average performance of groups voting under unanimity exceeds the performance of individuals and groups using majority. In our static treatments, the average performance of groups using either majority or unanimity exceeds that of individuals. These comparisons remain the same, if slightly less pronounced, when focusing on the last 15 session rounds, see the Online Appendix for details.

The performance measure assessed above necessarily inherits the randomness induced by the particular information processes participants face. Consider, for example, our static treatments. Given the choice of time spent on information collection, the resulting posterior depends on the realized sample path. This inherent randomness introduces noise in our assessments, which could render comparisons between treatments insignificant. Instead, one could consider *expected* welfare, accounting for the *expected* posterior implied by each choice of stopping times. Similarly, in our dynamic treatments, it is natural to consider the *expected* time induced by any choice of decision accuracy and assess performance accordingly.

In the dynamic case, for any threshold posterior p, the expected stopping time is $\mathbb{E}[t|p] = \frac{(2p-1)\log\left(\frac{p}{1-p}\right)}{\mu}$. Thus, we define the expected performance for the dynamic treatments as[41]

$$\lambda_{i,g}^{expected} = p_{i,g} - 0.2 \cdot \mathbb{E}[t|p_{i,g}]$$

In the static case, for any chosen duration t, the expected posterior is $\mathbb{E}\left[p|t\right] = \frac{1}{2}\left(\text{erf}\left(\frac{\sqrt{1.4t}}{2}\right) + 1\right)$. Accordingly, we define the expected performance for the static treatments as

$$\lambda_{i,g}^{expected} = \mathbb{E}[p|t_{i,g}] - 0.2 \cdot t_{i,g}$$

We report the estimated mean of these performance measures under $\lambda^{expected}$ and *Performance Level* in Table 5 (third column). As was the case with our benchmark performance measures, all dynamic treatments outperform all static treatments at any conventional significance level. Now, however, the noise reduction brings forth differences within the dynamic and static treatments. The dynamic unanimity treatment outperforms the dynamic individual and majority treatments ($p < 0.01$), while the static individual treatment is outperformed by static majority ($p < 0.01$) as well as static unanimity ($p < 0.05$). Thus, after accounting for noise inherited by our processes, we find statistically significant differences between and within the dynamic and static treatments (which remain when considering the last 15 rounds). Importantly, the efficacy of rules governing decision making depends crucially on the information-collection format in place.

Relative Performance We now assess how close participants came to the theoretically optimal performance. To do so, we calculate the theoretically optimal performance by utilizing the optimal

[41] As a caveat, in our dynamic treatments, we effectively assume time-independent thresholds for these estimates. This approximation simplifies assessments dramatically and yields results that are in line with those from our alternative performance measures.

posterior $p^* = 0.81$ for the dynamic case and the optimal wait time $t^* = 29.58$ for the static case.

$$\lambda^*_{dynamic} = p^* - 0.2 \cdot \mathbb{E}[t|p^*] = 0.68 \qquad \lambda^*_{static} = \mathbb{E}[p|t^*] - 0.2 \cdot t^* = 0.62.$$

In contrast, an immediate decision would yield an accuracy of 0.5 at no cost. This constitutes a plausible lower bound on performance and results in an expected payoff of $\underline{\lambda} = 0.5$.[42] A measure capturing the *relative* performance is then

$$\tilde{\lambda}_{i,g} = \frac{\lambda_{i,g} - \underline{\lambda}}{\lambda^*_j - \underline{\lambda}} \qquad\qquad j \in \{dynamic,\ static\}.$$

An estimated relative performance of 0 indicates that, on average, the treatment performs no better than an immediate decision that incorporates no information. In contrast, an estimated relative performance of 1 indicates the treatment exhibits optimal performance.[43] We report these estimated relative performances in Table 5 under *Relative Performance* (second and fourth columns).

As can be seen, static treatments exceed the dynamic treatments when it comes to information utilization. For example, focusing on our arguably less noisy measure of expected performance (fourth column), the dynamic treatments reach scores between 0.82 and 0.89, whereas the static treatments reach scores between 0.88 and 0.94 (comparisons that, again, remain virtually identical when focusing on the last 15 rounds alone).

Substantively, while information is utilized more efficiently in our static treatments, dynamic information collection is sufficiently more effective that it still yields greater expected welfare.

8 Conclusions

This paper reports results from a set of experimental treatments testing static and sequential sampling, in individuals and groups. Sequential sampling yields superior outcomes to static sampling, with groups under unanimity delivering the best outcomes. Contrasting theoretical predictions, sequential sampling yields time-decreasing thresholds and static sampling yield excessive information collection. Furthermore, groups behave differently than individuals, beyond mechanical aggregation effects. Majority rule yields the quickest decisions, particularly in sequential sampling.

Our experimental paradigm and our results point to several possible future directions of inquiry.

[42] It is certainly possible to achieve lower performance. For example, an excessively long wait can yield negative expected payoffs. We do not observe such behavior in our data.

[43] By definition, the comparison of decision-making rules within either our static or dynamic treatments coincides with the comparison generated by our expected welfare measure.

In our study of individuals, it would be interesting to consider behavior under a richer set of parameters. In our study of groups, we have focused on a baseline case in which the model predicts no group effects. This is intended as an initial benchmark on which to build a richer understanding of heterogeneous committees, as in the model studied in Chan et al. (2018). Specifically, similar experiments could be designed with group members experiencing heterogeneous preferences over alternatives and heterogeneous information costs. It would also be interesting to vary other features of groups: their size, the monitoring available to group members, etc.[44]

9 Appendix

9.1 Pulling the Trigger: Majority and Unanimity

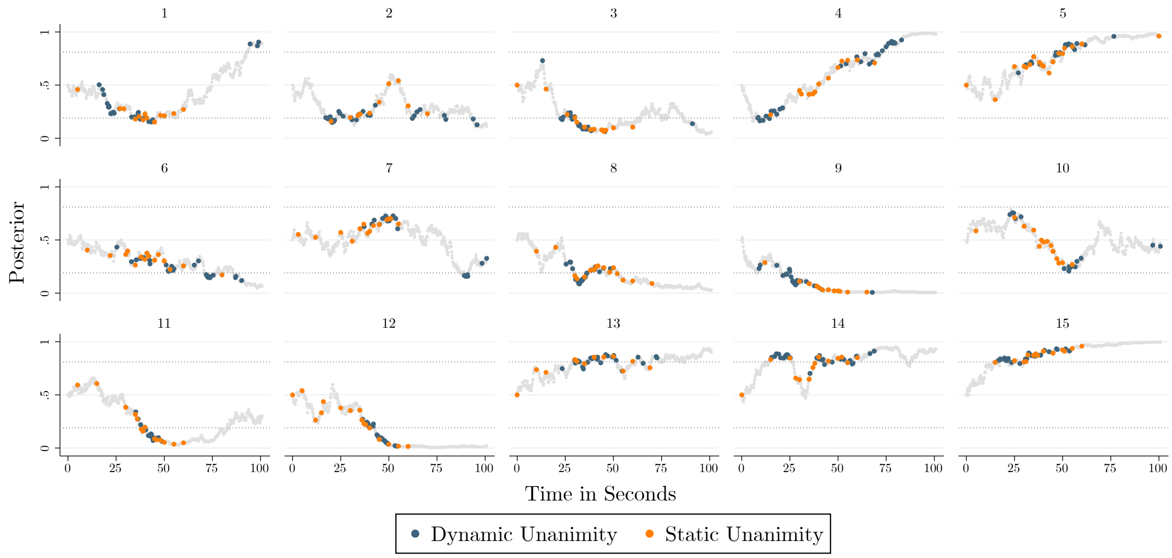
Figure 1 in the text depicts the evolution of posteriors and the corresponding choices in our individual treatments. Figure 10 and Figure 11 below provide analogous graphs for our majority and unanimity treatments, which depict decision posteriors corresponding to pivotal votes.

The general patterns observed for our individual treatments remain. For example, later decisions in our dynamic treatments often correspond to lower accuracies in our dynamic treatments and posteriors generated in our static treatments are more dispersed than those emerging from our dynamic treatments. 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 10: Pulling the Trigger: Majority Treatments

[44] Naturally, the study of larger groups would entail some non-trivial logistical hurdles as most physical laboratories are limited in size.

Figure 11: Pulling the Trigger: Unanimity Treatments



9.2 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 6 displays regression analysis pertaining to individual and group choices—the stopping posterior—in our dynamic 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 6 echoes our observations in the text. We see significant differences 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

Table 6: 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

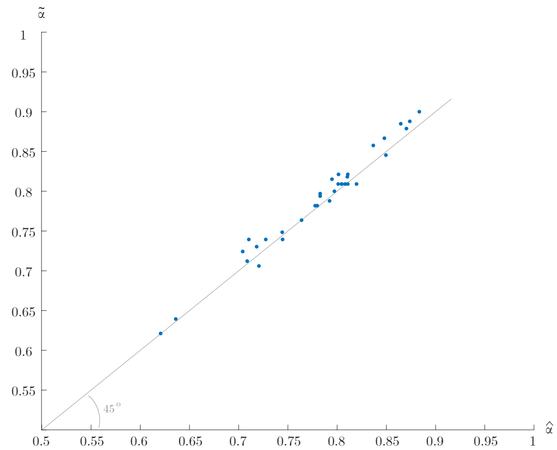
Standard errors in parentheses
Individual-level clustering
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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 6 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.[45] The estimated coefficients corresponding to decision times

[45]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 12: 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.

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.

9.3 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 parameter values are then $\tilde{\alpha}$, $\tilde{\beta}$, and $\tilde{\sigma}_\varepsilon$ that minimize the sum of squared errors.[46]

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

Table 7: Dynamic Treatment Learning

	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

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

Figure 12 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.

9.4 Learning

Dynamic Treatment Learning 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 7 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 9.2 above); and an interaction between *Round* and *Slow*, allowing for a different learning trend depending on the process.[47] 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 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

[47] 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 8: Static Treatment Learning

	Time Waited								
	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	-0.223*	-0.758***	-0.140	-0.260***	-0.628***	-0.0220	-0.444***	-0.767***	-0.241*
	(0.117)	(0.232)	(0.138)	(0.0931)	(0.211)	(0.130)	(0.0900)	(0.190)	(0.128)
$Correct_{t-1}$	-1.964*	-2.228*	-0.0829	-1.458**	-1.856	-0.747	0.0774	-0.574	0.619
	(1.020)	(1.093)	(1.437)	(0.623)	(1.369)	(0.716)	(0.813)	(0.901)	(0.930)
$Difference_{t-1}$				0.384**	0.381**	0.0656*	0.140***	0.0669	0.0685
				(0.155)	(0.165)	(0.0366)	(0.0456)	(0.0596)	(0.0562)
Individual Level FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	899	434	465	1392	672	720	1305	630	675

Standard errors in parentheses

Individual-level clustering

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

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.

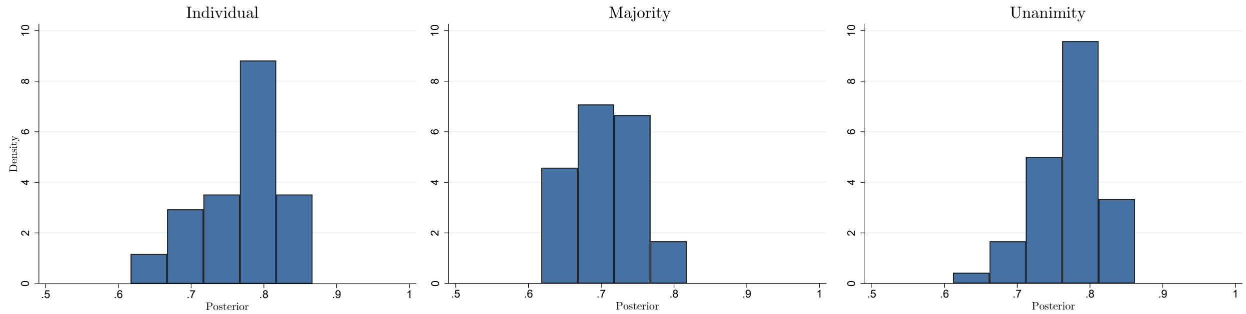
Static Treatment Learning We now perform a similar analysis for the static treatment. The specification of the regressions presented in Table 8 is as described in Section 9.4. However, the dependent variable, corresponding to participants' choice, is now desired duration rather than stopping posterior. Furthermore, in the static case, participants cannot react differently to slow and quick sample paths, since those evolve only after decisions have been made. Thus, we do not include *Slow* and *Round* × *Slow* in the regressions of Table 8.

The estimated coefficients of $Correct_{t-1}$ and $Difference_{t-1}$ are typically insignificant: participants do not greatly react to whether or not the last-period decision was correct, or to the difference between their last-period decision and other group members' last-period decision. Once more, the magnitude and statistical significance of *Round* greatly diminishes in the last 15 rounds. This leads us to believe that participants had sufficient rounds to learn and adjust their decisions.

9.5 Individual-Level Choice Heterogeneity

Figure 13 depicts the distributions of the mean posteriors at the time of voting in our dynamic treatments, where mean posteriors are calculated for each participant separately.

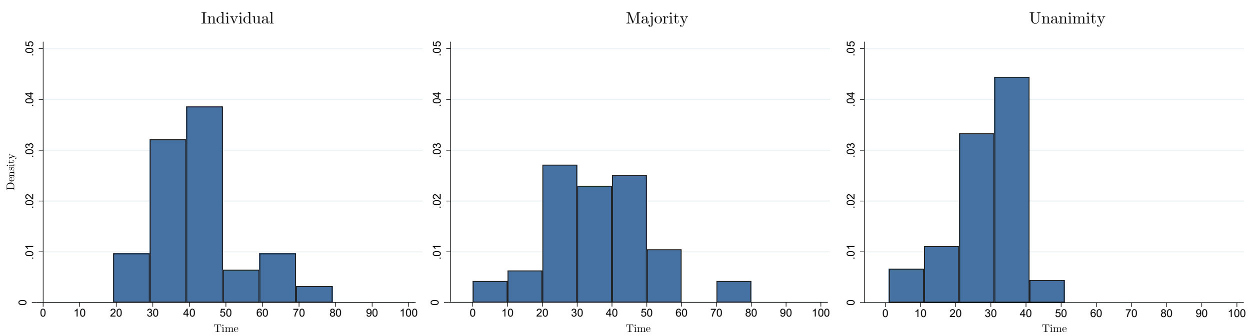
Figure 13: Dynamic Treatments Individual Heterogeneity



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.

Figure 14 shows the distribution of the mean desired times, calculated separately for each participant, in our static treatments.

Figure 14: Static Treatments Individual Heterogeneity



In the static treatments, we also see somewhat narrower distributions of behaviors under the individual and unanimity treatments than in the majority treatment. However, in line with our discussion, these distributions are concentrated at different desired wait durations across our treatments.

References

Ambuehl, S. and Li, S. (2018). Belief updating and the demand for information. *Games and Economic Behavior*, 109:21–39.

Baldassi, C., Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M., and Pirazzini, M. (2020). A behavioral characterization of the drift diffusion model and its multialternative extension for choice under time pressure. *Management Science*, 66(11):4921–5484.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, 82(6):2005–2039.

Becker, G. M. (1958). Sequential decision making: Wald's model and estimates of parameters. *Journal of Experimental Psychology*, 55(6):628.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.

Bhattacharya, S., Duffy, J., and Kim, S. (2017). Voting with endogenous information acquisition: Experimental evidence. *Games and Economic Behavior*, 102:316–338.

Brown, M., Flinn, C. J., and Schotter, A. (2011). Real-time search in the laboratory and the market. *American Economic Review*, 101(2):948–74.

Canen, N. (2017). Information accumulation and the timing of voting decisions. *mimeo*.

Caplin, A., Dean, M., and Martin, D. (2011). Search and satisficing. *American Economic Review*, 101(7):2899–2922.

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

Chapman, J., Snowberg, E., Wang, S., and Camerer, C. (2019). Loss attitudes in the us population: Evidence from dynamically optimized sequential experimentation (dose). *mimeo*.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chen, S. and Heese, C. (2021). Fishing for good news: Motivated information acquisition. *mimeo*.

Cooper, D. J. and Kagel, J. H. (2016). Other-regarding preferences. *The Handbook of Experimental Economics*, 2.

Darley, J. M. and Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, 8(4p1):377.

Dominitz, J. and Manski, C. F. (2017). More data or better data? a statistical decision problem. *The Review of Economic Studies*, 84(4):1583–1605.

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.

Edwards, W. and Slovic, P. (1965). Seeking information to reduce the risk of decisions. *The American Journal of Psychology*, 78(2):188–197.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–38.

El-Gamal, M. A. and Palfrey, T. R. (1996). Economical experiments: Bayesian efficient experimental design. *International Journal of Game Theory*, 25(4):495–517.

Elbittar, A., Gomberg, A., Martinelli, C., and Palfrey, T. R. (2020). Ignorance and bias in collective decisions. *Journal of Economic Behavior & Organization*, 174:332–359.

Fehr, E., Herz, H., and Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review*, 103(4):1325–59.

Fischer, P., Jonas, E., Frey, D., and Schulz-Hardt, S. (2005). Selective exposure to information: The impact of information limits. *European Journal of Social Psychology*, 35(4):469–492.

Fried, L. S. and Peterson, C. R. (1969). Information seeking: Optional versus fixed stopping. *Journal of Experimental Psychology*, 80(3p1):525.

Fudenberg, D., Strack, P., and Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–3684.

Gabaix, X., Laibson, D., Moloche, G., and Weinberg, S. (2006). Costly information acquisition: Experimental analysis of a boundedly rational model. *American Economic Review*, 96(4):1043–1068.

Gerardi, D. and Yariv, L. (2007). Deliberative voting. *Journal of Economic Theory*, 134(1):317–338.

Gerardi, D. and Yariv, L. (2008). Information acquisition in committees. *Games and Economic Behavior*, 62(2):436–459.

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.

Green, P. E., Halbert, M. H., and Minas, J. S. (1964). An experiment in information buying. *Journal of Advertising Research*, 4(3):17–23.

Greene, W. H. (2018). *Econometric Analysis, 8th Edition*. Pearson/Prentice Hall.

Gretschko, V. and Rajko, A. (2015). Excess information acquisition in auctions. *Experimental Economics*, 18(3):335–355.

Henry, E. and Ottaviani, M. (2019). Research and the approval process: The organization of persuasion. *American Economic Review*, 109(3):911–55.

Hoffman, M. (2016). How is information valued? evidence from framed field experiments. *The Economic Journal*, 126(595):1884–1911.

Imai, T. and Camerer, C. F. (2018). Estimating time preferences from budget set choices using optimal adaptive design. *mimeo*.

Kagel, J. H. and Roth, A. E. (1995). *The Handbook of Experimental Economics*. Princeton university press.

Luce, R. D. et al. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press on Demand.

Martinelli, C. (2006). Would rational voters acquire costly information? *Journal of Economic Theory*, 129(1):225–251.

McClellan, A. (2021). Experimentation and approval mechanisms. *mimeo*.

Neyman, J. and Pearson, E. S. (1933). Ix. on the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337.

Persico, N. (2004). Committee design with endogenous information. *Review of Economic Studies*, 71(1):165–191.

Pikulina, E. S. and Tergiman, C. (2020). Preferences for power. *Journal of Public Economics*, 185:104173.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922.

Ratcliff, R. and Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2):333.

Stephan, F. F. (1948). History of the uses of modern sampling procedures. *Journal of the American Statistical Association*, 43(241):12–39.

Strack, P. and Viefers, P. (2021). Too proud to stop: Regret in dynamic decisions. *Journal of the European Economic Association*, 19(1):165–199.

Strulovici, B. (2010). Learning while voting: Determinants of collective experimentation. *Econometrica*, 78(3):933–971.

Swensson, R. G. (1972). The elusive tradeoff: Speed vs accuracy in visual discrimination tasks. *Perception & Psychophysics*, 12(1):16–32.

Szkup, M. and Trevino, I. (2021). Information acquisition and self-selection in coordination games. *mimeo*.

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

Wald, A. (1947). *Sequential Analysis*. New York.