Individual and Collective Information Acquisition: An Experimental Study*

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

September 2020

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. Generally, outcomes approximate the theoretical benchmark and sequential information collection is welfare enhancing relative to static collection. Nonetheless, several important departures 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.

*We thank Roland Benabou, Stephen Morris, Salvo 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, Princeton University preshidi@princeton.edu

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

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

¶Department of Economics, The Chinese University of Hong Kong jimmyhingchan@cuhk.edu.hk

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

1 Introduction

1.1 Overview

Juries, boards of directors, congressional and university committees, government agencies such as the FDA or the EPA, and many other committees spend time deliberating issues before reaching a decision or issuing a recommendation. An important component of such collective decisions is the acquisition of information. The statistics literature has offered two leading models of costly information collection. Perhaps the most well known and heavily utilized is classical hypothesis testing, where the amount of information to be collected, often the size of a data set, is chosen at the outset. Classical hypothesis testing has many practical advantages. Most notably, it requires only one choice pertaining to the information volume, or sample size, to be collected. Sequential hypothesis testing, going back to Wald (1947), calls for incremental choices of information collection. The researcher sees one piece of information, then decides whether to proceed with another, and so on. Sequential hypothesis testing has efficiency advantages: information collection occurs only when its marginal benefits justify its cost. However, it is arguably more complex, requiring repeat decisions and information monitoring over time.

In this paper we provide an experimental examination of how individuals and groups collect information. We examine both static and sequential information collection by both individuals and groups following a variety of decision-making protocols.

The main results of our investigation are the following. First, although average participants' behavior is arguably close to the theoretical predictions, we see several consistent deviations. In particular, we observe excessive information collection when information collection is static, as in the classical hypothesis testing model. We also see agents' becoming less demanding of accuracy over time when information collection is sequential. Second, individuals and groups behave markedly differently. Furthermore, the collective rules by which groups make decisions have substantial impact on outcomes. Specifically, groups making decisions under majority rule make far hastier decisions, utilizing substantially less information, than either individuals, or groups that decide under unanimity rule. Ultimately, we see similar decision accuracies under static or sequential protocols. Nonetheless, when accounting for information costs, sequential protocols yield greater efficiency levels.

The investigation of information collection, and deliberative processes more generally, is particularly challenging using field data. It is often difficult to assess the precision of samples collected and the underlying preferences of decision makers. Natural procedures, such as those pertaining to juries, boards of directors, or the FDA, often have rigid protocols and are therefore difficult to compare in a controlled fashion. Lab experiments are therefore particularly useful in these settings.

At the core of our experimental design is the following decision problem. There are two ex-ante equally likely states, A or B—a metaphor for a guilty or innocent defendant, an investment that is worthwhile or not, etc. Ultimately, each participant needs to 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, the cost of observing the realized Brownian path. 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 non-trivial 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 static setting emulates the classic hypothesis testing scenario. Participants determine, at the outset, the time horizon during which they collect information—namely, observe the Brownian path. They then indeed see the path unravel for the specified amount of time, get informed of the ultimate posterior over states, and make their guess. In the dynamic setting, emulating the sequential sampling scenario, participants track the evolution of the Brownian path and can stop at any time to submit their guess.

As our introductory examples suggest, there are many applications in which information collection is undertaken by a committee. This motivates our choice of treatments. In some treatments, decisions are made by individuals, as in the classic paradigms. In others, they are made in groups. When in a group, we consider two commonly-used institutions: majority and unanimity. 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 the dynamic setting, group members decide at each point in

time whether to stop or continue searching. Under majority, whenever two members wish to stop and agree on a guess, information collection terminates for the group, and the majority guess is submitted. Analogously, under unanimity, whenever all members wish to stop and agree on the guess, information collection terminates and that guess is implemented. In particular, in all our group treatments, all group members receive the same payoff, derived from the jointly-determined information cost, and the guess' accuracy.

Our individual treatments offer a natural benchmark for the basic predictions emerging from the canonical statistical information-collection procedures. In the static setting, our parameters are such that the optimal information-collection horizon is 30 seconds. In our experiments, individuals choose 42 seconds, a choice that is 40% higher than is optimal. In the dynamic setting, it is optimal to use a constant threshold on posterior beliefs, set at 0.81 for our parameters. Intuitively, whenever one becomes sufficiently confident in the assessment of which state had been realized, the cost of further information collection outweighs its benefits. In our experimental treatments, individuals' mean posteriors at decision time is remarkably similar to that predicted by theory, standing at 0.77. Nonetheless, individuals do not seem to use constant thresholds. We see decreasing threshold over time, with participants becoming more lenient as time passes.

By design, our groups are homogeneous. Theoretically, there is a unique efficient equilibrium mimicking the optimal individual choices. Therefore, our group treatments allow for the investigation of pure group effects. We find that groups behave differently from individuals, and that this behavior depends on the voting rule governing group decisions.

Majority and unanimity generate different behaviors and outcomes in our setting. Groups governed by majority decide much faster than individuals, and therefore under-collect information to an even greater extent. Groups governed by unanimity decide more slowly than individuals, and come extremely close to the theoretical benchmark in terms of decision accuracy.

Individuals choosing on their own exhibit heterogeneity in behavior. Could the mere grouping of heterogeneous individuals explain the patterns observed in our group treatments? In order to answer this question, we simulate groups composed of participants from our individual treatments and record how such artificial groups would have decided under majority and under unanimity, absent any changes in behavior. Differences between these simulated groups' outcomes and individuals' capture a mechanical effect of aggregating heterogeneous individuals.

We find that differences between outcomes of groups using unanimity and individual outcomes can be fully explained through the mechanical effect of aggregation.[1] In contrast, majority decides substantially faster than simulated groups of heterogeneous individuals. This presents a puzzle: why are groups deciding using majority rule so hasty, while groups deciding using unanimity are not? We suggest that majority creates a *demand for agency* that leads agents to vote early. Remarkably, we find that the decision accuracy under majority replicates the decision accuracy of the most lenient members in our simulated groups. This feature suggests an additional force impacting the first voter under majority. That first voter can accelerate her vote in order to strategically influence the timing of the second, pivotal vote.

We then turn to a comparison between static and sequential information collection. We find that, consistent with theoretical predictions, sequential information collection outperforms static information collection. However, often decision bodies' decisions affect a large segment of the population. The decision's accuracy is then of much greater import than then cost experienced by a small fraction of society. When considering decision accuracies, sequential information collection no longer dominates static information collection. In fact, under majority rule, static information collection leads to superior accuracy relative to sequential information collection.

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 18'th century. Classical hypothesis testing has been used, formally or informally, for centuries, see Stephan (1948). It came of age with the development of statistical hypotheses tests by Neyman and Pearson (1933), who showed that the likelihood ratio test is the most powerful test for assessing hypotheses with a given data set. Examples abound for its uses. It is still arguably the most heavily applied approach for deducing inferences from limited observation sets. 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

[1] The effect is nonetheless real: every individual in a group within our unanimity treatment is affected by the group member with the most stringent threshold.

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 the 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 whether to 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 (2017) build on the literature on sequential hypothesis testing that started with Wald (1947).

In terms of experiments, 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 useful information underreacts to increased informativeness and that individuals value information that may yield certainty disproportionately 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 information when signals

are valuable and overpay when signals are less valuable. Our static treatments add to this literature by illustrating how both individuals and groups resolve the accuracy-cost trade-off.[2]

To our knowledge, there is little experimental work that speaks directly to the sequential sampling setup.[3] 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. However, the underlying optimization problem is quite different from ours.

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 insight 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).

2 Experimental Design

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 neutrally labeled 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 financial context, etc. At the start of each period, one of the states is chosen at random with equal probabilities. Participants ultimately need to guess which state had been chosen and are paid according to whether or not their

[2] Several studies inspect information collection in strategic settings different from ours. See for example, Elbittar et al. (2016) and Bhattacharya et al. (2013), who consider information aggregation settings in which individuals acquire private information, Szkup and Trevino (2015), who explore information collection in the context of global games, or Gretschko and Rajko (2015), who focus on auctions.

[3] Interestingly, the idea of using sequential experimental *designs* has been suggested in various contexts, see El-Gamal and Palfrey (1996), Chapman et al. (2018), Imai and Camerer (2018), and references therein.

guesses are correct. In the lab, participants receive $2 for a correct guess and nothing otherwise.

Prior to making a choice, participants have access to information that evolves according to a continuous-time Weiner process. When the state is A, the process has drift μ and variance σ^2; When the state is B, the process has drift $-\mu$ and variance σ^2. Throughout our treatments, $\mu = 0.84$ and $\sigma^2 = 1$

There are two dimensions that we vary across our treatments: whether information acquisition decisions are static or sequential and whether choices are made by individuals, groups using majority rule, or groups using unanimity rule.

In what follows, we begin by describing our sequential treatments, which are the more novel part of our experiment. The design choices of the dynamic treatment also guided our choices for the design of the static treatments, which are described next.

Sequential Sampling In our *dynamic* treatments, participants observe the information evolve over time and, at each instant, have a choice of guessing A, B, or waiting for further information by choosing W. Time spent acquiring information comes at a fixed cost of 40 cents a minute.

In the treatment in which individuals make decisions on their own—the Individual Dynamic treatment—a round ends for a participant as soon as he or she selects one of the A or B guesses.

In our group treatments, participants are randomly matched to create groups of 3 in each round. A round ends as soon as a quorum of q individuals agrees on an A or B guess. In the Majority Dynamic treatment, $q = 2$, whereas 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 observe choices others made within their group.

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 would like to spend collecting information. As in the dynamic treatment, information costs are fixed at 40 cents a minute.

When individuals make decisions independently—the Individual Static treatment—they observe the information evolve for their desired time.[4] Their guess is then automated to reflect the state

[4]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.

that is more likely given the information collected: either A or B.[5]

Our static-sampling group treatments are analogous to those corresponding to the dynamic treatment. In each round, participants are matched into groups of 3. At the outset of each round, participants submit simultaneously their desired waiting time. The resulting group waiting time is the median desired waiting time of group members in our Majority Static treatment; it is the maximal desired waiting time of group members in our Unanimity Static treatment. As in the individual variant, participants observe the information evolve for the group's waiting time. The group guess, A or B is again automated to best respond to the information collected.

Feedback and Payments In all treatments, the feedback at the end of each round contains participants' payoffs and other group members' choices whenever relevant.

Each treatment was preceded by two practice rounds, followed by 30 "real" rounds. Participants were ultimately paid for 20 randomly selected rounds.

Information Processes The 30 information processes experienced by participants were identical across treatments. These processes were selected in the following way. We randomly generated 15 Weiner processes, with the parameters specified above, that are "representative" in that the mean, median, and five quintiles of the theoretically optimal sequential stopping times matched 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 was A (or B), it was B (or A) in the reflected processes. Furthermore, at any time t, if the original process indicated a probability p that the state is A, the reflected process indicated 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, with a gap of 15 rounds in between. This design element allows us to evaluate learning in a highly controlled fashion.[6]

[5]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. Note that this choice could not easily be automated in the dynamic treatment.

[6]Because we describe the evolution of a process over time through posterior probabilities that change over time (on our interface, five times a second), it is practically impossible for subjects to identify these effective process

The evolution of a Weiner process continuously provides information on the likelihood of either state prevailing. Nonetheless, the Bayesian calculus necessary to deduce this likelihood is non-trivial and this difficulty is orthogonal to our investigation. Indeed, it is well known that lab participants are frequently challenged by statistical updating, see references in our literature review. In order to mitigate the impacts of subjects' limitations exclusively pertaining to statistical analysis, in our design, participants are presented with the evolution of the *probability* that the state is A directly.

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 mean higher than 1 and non-trivial variance (e.g., one paying 2.5 the amount invested with probability 50%). In addition, participants participated 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.[7]

Summary The experiments were run at the Princeton Experimental Laboratory for the Social Sciences (PExL) with 254 participants. Each treatment entailed 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. The experimental software was programmed using oTree (Chen et al., 2016).

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

repetitions. Such recollection would require the memory of hundreds of ordered values and the realization that they are mirrored.

[7]Elicitations were duplicated in order to allow for measurement-error correction as suggested in Gillen et al. (2019).

3 Theoretical Predictions

We now briefly discuss the optimal information-collection policies in our various treatments. For details, see Dvoretzky et al. (1953) or Chan et al. (2018).

We assume a setting as 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 Weiner process with a variance of 1. When the state is A, the process has drift $\mu = 0.84$; When the state is B, the process has drift $-\mu = -0.84$ and variance 1. Tracking this information comes at a flow cost of c. We assume the agent guesses the state that is more likely once information collection terminates. For ease of presentation, 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 denote by $\mu' \equiv 2\mu^2$. The agent's posterior belief is then given by a Wiener process, with drift μ' and instantaneous variance $2\mu'$ under state A and drift $-\mu'$ and variance $2\mu'$ under state B. A higher value of μ' (higher μ or lower ρ) indicates a more informative process. Given our parameters, $\mu' = 1.4$.

3.1 Static Treatments

The probability of guessing the true state correctly at any given time t is:

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

In the static setting, a risk-neutral agent maximizes:

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

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 or product log function.

With our experimental parameters, $t^* = 0.49$, or 29.58 seconds, since one unit of time in the

lab is one minute. Thus, in expectation, a risk-neutral agent maximizes her payoff by waiting for 29.58 seconds.[8]

Consider now a group of $n > 1$ identical 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 state realized when information collection terminates. 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 a best response for any agent, regardless of the strategies other agents in the group utilize.

3.2 Sequential Treatments

One of the main contribution of Wald (1945) and the continuous-time counterpart of Dvoretzky et al. (1953) is to illustrate 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 $\theta_t = \log\left(\Pr[A]/\Pr[B]\right)$. 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:[9]

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

The corresponding first-order condition with respect to the lower boundary is then:[10]

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

[8] Analysis of this setting in the presence of risk aversion is presented in Section 11.1 of the Appendix. This analysis suggests that risk aversion has no substantial impact on behavior.

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

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

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

Because the problem is symmetric, the solution satisfies $g = -G$. The optimal value of G can then be determined by:

$$c\left(2e^{G}G + e^{2G} - 1\right) - e^{G}\mu' = 0$$

With $\mu' = 1.4$ and $c = 0.2$, we numerically calculate the optimal boundary as $G^* = 1.461$. Translated into probabilities, this value becomes $\frac{e^{1.461}}{1+e^{1.461}} = 0.81$. Thus, in the dynamic version, the theoretical prediction is that a risk-neutral agent should wait until the probability of the most likely state is 81%.

Consider now a group of $n > 1$ identical agents. At each date, 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 corresponds to the optimal search policy described above, namely utilizing a threshold of 81%. Furthermore, as long as agents use symmetric cutoff policies, this choice is a best response for any agent, regardless of the cutoffs chosen by other agents in the group.

4 Approach to Data Analysis

As may be expected, subjects behavior changes during the early rounds as they learn about the problem. However, most of the learning that we observe occurs within the first 15 rounds. In fact, we see no evidence for substantial learning at later rounds. For details, see Section 11.4 in the Appendix. Throughout the paper, we present figures aggregated across all experimental rounds as those displayed appear virtually identical when we use either the first half or the second half of our sessions. Regression results are presented for data corresponding to all rounds and to the last 15 rounds. Recall that, in our design, the first and last 15 rounds utilized the same ordered set of information processes. Thus, the sample of settings participants encounter in the first and second

half of each session is identical.

Risk attitudes and altruism proclivities do not appear to explain any aspect of our data, even after measurement-error correction. We therefore do not include data from these elicitations in our main specifications. Section 11.3 in the Appendix offers some additional analyses that explicitly speak to this claim.

5 Broad Patterns of Behavior

Table 2 displays an aggregate overview of some of our results. It illustrates the mean posterior when a decision has been made and the mean time for a decision across our treatments. 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 statistically indistinguishable from those theory predicts. Furthermore, the Majority Dynamic treatment corresponds to the least amount of waiting, an observation we shall return to.

Differences between observed decision posteriors and those predicted by theory may, at first blush, appear small. Nonetheless, these differences translate to large differences in wait times. For instance, the Unanimity Dynamic treatment leads to double the wait time compared to the Majority Dynamic treatment. This is a common feature in information-collection settings, where the cost of precision is effectively convex—the higher is the current posterior precision, the more time needs to be spent to establish a certain marginal precision increase.

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.

When comparing the static and dynamic treatments, we see that, contrary to the theoretical predictions, mean decision times are longer in the static treatments. Furthermore, mean posteriors at decision times are comparable or only slightly lower than those observed in our dynamic treatments, which is also in contrast with theoretical predictions. These observations have clear welfare implications. When committees collecting information make decisions that affect a large population, such as juries, the FDA, and so on, the population welfare, captured by the quality of decisions, is similar under both static and dynamic protocols. We return to this point when discussing performance in our different treatments in Section 8.

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	

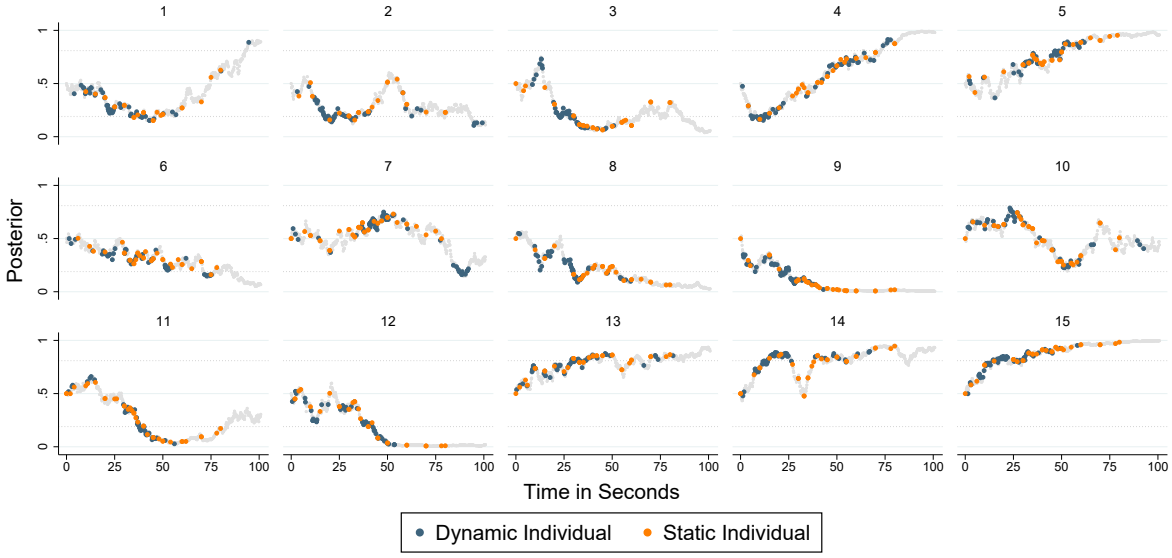
Standard errors in parentheses

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."

The Figure suggests some important themes that appear in our more detailed analysis below. First, it is apparent that decisions are quite heterogeneous, with some individuals demanding a lot more accuracy than others. Second, many observations are close to optimal. In particular, in the dynamic setting, participants clearly respond to information in that decisions are more highly clustered around higher posteriors. Moreover, many decisions are taken with accuracy predicted by theory (and corresponding to the horizontal dashed lines within each panel). Third, individuals in the dynamic setting appear to be more lenient over time, requiring less accuracy to stop. Consider, for example, process 10. Several individuals decide late in the process, when posteriors are close to 50%, despite choosing not to stop at two 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. Processes 2 and 14 provide extreme examples. In these processes, some static decisions involve a substantial wait time, but culminate in decision posteriors of around 50%. Continuation would clearly be preferable if agents could see the posterior (as they do in the dynamic setting). In contrast, in processes 12 and 15, some static choices take place at extremely high posteriors. Earlier stopping could have been preferable if agents had been able to condition their behavior on the history.

The analogous figure for our majority and unanimity treatments appears in Section 11.2 in the Appendix. As we soon discuss, behavior is different in those treatments and it is natural to compare only pivotal agents under the two decision protocols, majority and unanimity. Nonetheless, some observations remain. We see heterogeneous decisions, responses to information and more leniency over time in the dynamic treatments, and decisions at extreme posteriors, either low or high, in the static treatments.

Figure 1: Pulling the Trigger: Individual 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.

$std(p_{dynamic}) = 0.990$, $std(p_{static}) = 0.134$, $std(t_{dynamic}) = 23.22$, $std(t_{static}) = 15.16$

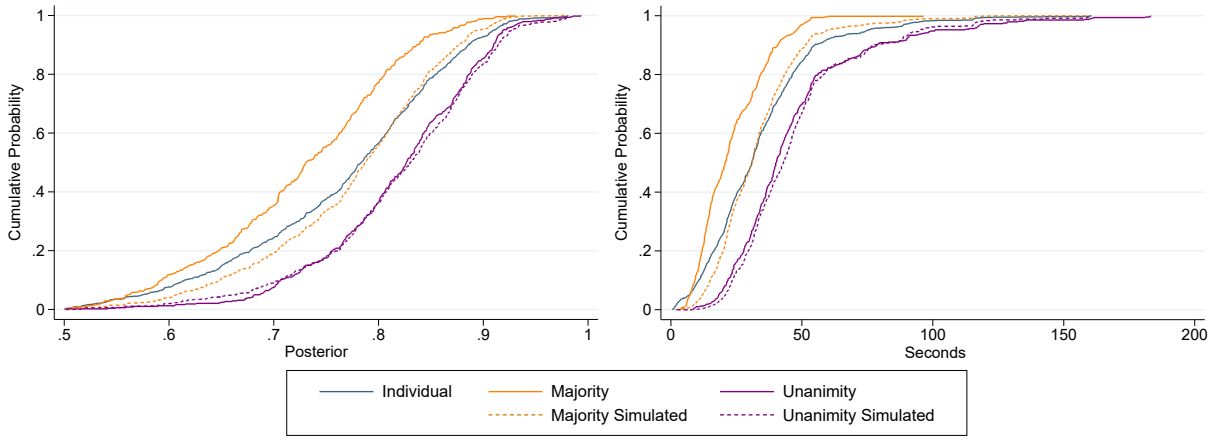
6 Sequential Information Collection

6.1 The Impacts of Decision Procedures

We consider three procedures for information collection and choices: by individuals, by groups using majority, and by groups using unanimity. Figure 2 displays the cumulative distribution functions of decision posteriors on the left panel, as well as induced times on the right panel, for each of

these procedures. We see an important impact of the governing decision rule. Distributions are 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.

Figure 2: Dynamic Treatment CDFs



Given the heterogeneity we observe at the individual level, one may naturally wonder whether the differences we observe across our dynamic treatments are simply a mechanical artifact of the grouping of three random individuals that respond heterogeneously to the task at hand. In order to assess whether the differences we see among these treatments are purely mechanical, we artificially generate groups of three from our individual treatment.[11]

Figure 2 presents the resulting cumulative distribution functions from these simulated groups in addition to the distributions we observe in our data. As can be seen, the additional accuracy granted by groups using unanimity appears to be a purely mechanical phenomenon. In fact, a two-sided Kolmogrov-Smirnov (K-S) test, with a p-value of 0.345, fails to reject the hypothesis that these distributions, the simulated and observed unanimity group decisions, are identical.

In contrast, our simulated groups using majority rule yield substantially more accurate decisions than participants in the group majority treatment, suggesting that hasty majority choices are not a mechanical effect. A two-sided K-S test does not reject the hypothesis that these distributions

[11] Specifically, for each round, we randomly group the 34 participants in our individual treatment into 11 groups of 3 participants each a 1,000 times. Across all 30 rounds, 330,000 groups are then simulated.

are different.[12]

6.2 Declining Thresholds

Table 3 displays regression analysis pertaining to group and individual choices—the posterior at which the pivotal vote was cast—in our dynamic treatments. We use the short-hand 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, that indicate whether observations are taken from the last 15 rounds of our sessions. Last, we consider the impacts of time spent collecting information. 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 it is labeled as "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. [13]

The first column of Table 3 echoes our observations from the previous section. We see significant differences between treatments, with less precise, or hasty, majority decisions and more precise, or longer, unanimous decisions. Compared to the individual treatment, the mean posterior with which the pivotal majority vote is cast is about 4 percentage points lower, while the mean posterior with which the unanimity vote is cast is about 5 percentage points higher.

Throughout, we see a significant effect of learning over the first 15 rounds, with participants becoming more patient, casting their vote with a significantly higher decision posterior. Since both the individual and majority treatments choose, on average, at posteriors well below the theoretically optimal, the increase in decision posteriors reflects changes towards the optimal choice. In the unanimity treatment, however, learning leads to overshooting, with an average decision posterior

[12] Certainly, observations generating these figures are correlated. This makes standard statistical tests for comparing these distributions questionable. We soon use regression analysis to statistically determine what affects decisions in terms of both the institution in place and the governing information process in question.

[13] Alternative specifications are presented in Section 11.3.1 of the Appendix.

of 0.84 in the last 15 rounds. As mentioned at the outset, and elaborated on in the Appendix, 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, whether it is slow or quick. Slow processes are associated with significantly lower decision posteriors across our 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. Theoretically, whether a process is slow or quick should not impact the emergent decision posterior, only the time at which the decision is taken. Lower decision posteriors in slow processes indicate a non-stationary threshold for halting information collection. The observation hints at the idea that individuals and groups become less demanding of accuracy as time progresses.

The last two columns of Table 3 illustrate a declining-threshold pattern more directly. Namely, we introduce an explicit dependence on the time at which a pivotal vote is cast.[14] The estimated coefficients corresponding to decision times are negative and statistically significant: the longer it takes for the pivotal vote to be cast, the lower is the decision posterior. As before, the least affected treatment is unanimity and the most affected treatment is majority. In particular, in the last 15 rounds, since the estimated parameter value for *Time M* is -0.0018, for each 5 seconds that the group decision is delayed, the typical decision posterior decreases by about one percentage point.

Our finding that thresholds are decreasing is connected to the drift-diffusion model DDM (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 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). The explanation provided by Fudenberg et al. (2018) for the relationship between speed and accuracy relies on decision makers being uncertain about the process they face. In our setting, this is, in principle, not a relevant explanation since all features of the problem are known. Of course, it could

[14]The fixed-effect specification is appropriate since, without it, we could conceivably identify a misleading positive association between decision times and decision posteriors. Indeed, mechanically, since we consider a diffusion with drift, posteriors naturally 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.

be the case that participants experience subjective uncertainty of the type present in Fudenberg et al. (2018). However, learning does not significantly reduce the degree to which thresholds are decreasing in our setting. Since such "subjective errors" would come at a cost, we find this an unlikely explanation of our observations.[15]

Table 3: 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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

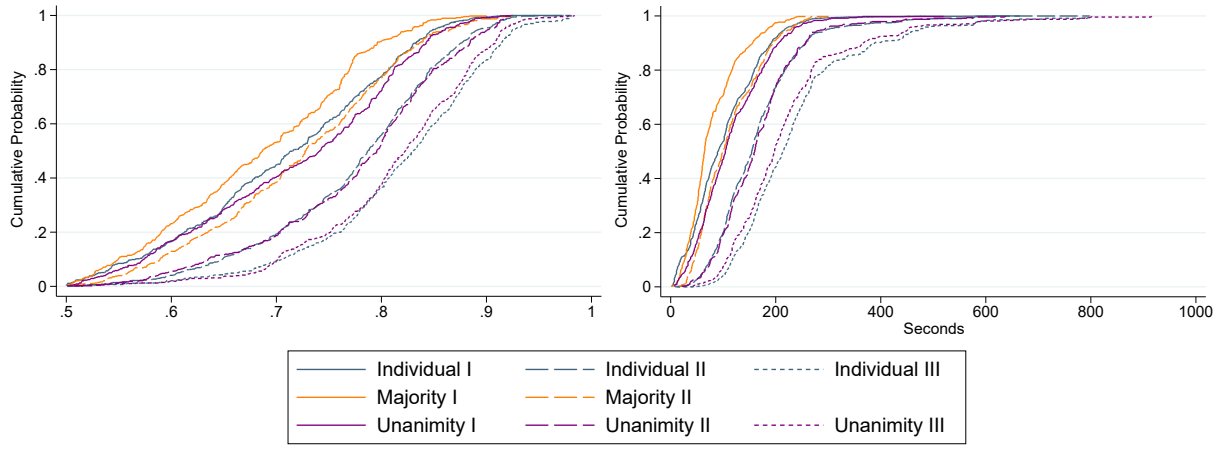
6.3 Voting First, Second, and Third

In our experimental interface, we record all votes cast by participants. In a group setting, if a pivotal vote has not been cast yet participants are allowed to change their minds. That is, they can change their vote, from say A, to W, or to B, and vice versa. Thus, it is not straightforward to determine how each vote corresponds to the first, second, and third order statistic. However, in the majority treatment 86% of games end the moment the second participant casts their first vote,

[15] Brown et al. (2011) provide an experimental analysis of sequential search. Because the model is stationary, the optimal reservation value is a constant wage. In analogy to our declining-threshold observation, their main finding is that participants' reservations values sharply decline over time. They attribute this phenomenon to participants experiencing non-stationary subjective costs of time spent searching.

whereas in the unanimity treatment 85% of games end the moment the third participant casts their first vote. Hence, given our data, using the first vote cast by each participant, we believe, is a good approximation of the first, second, and third order statistic in both the majority and the unanimity treatments. In Figure 3 we present the distribution of the first, and second vote in the majority treatment, and the distribution of the first, second, and third vote in the unanimity treatment. Alongside these distributions, we simulate groups of 3 generated from the individual treatment via the procedure described in Section 6 and present the resulting distribution of the first, second, and third order statistics.

Figure 3: Dynamic Treatment CDFs by Vote Order



An implication of our discussion in Section 6.1 is that 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. Figure 3 confirms this finding and reveals that this similarity also holds for the first and second order statistic/voter. This figure therefore reinforces the idea that individual voter behavior under unanimity is very similar to behavior of individuals when they are not in a group, and that the differences in outcomes under unanimity are exlusively due to the aggregation rule. Regarding the majority treatment, Figure 3 demonstrates that hasty behavior is not only a characteristic of the second (and pivotal) voter; the first voter appears to be quite hasty as well. Both the first and second order statistic from the simulated individual treatment stochastically dominate the first, and second voter 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.

6.4 Hasty Majority Decisions and a Demand for Agency

Why are majority decisions so hasty, whereas unanimity decisions are not? One possible explanation is a demand for agency. Prior work suggests that individuals have a taste for agency, that is, the ability to influence outcomes.[16] When operating as individuals, or as members of a group under unanimity voting, this agency is guaranteed because, 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 cast their votes. Thus, agency would elude a participant who is more demanding in terms of her threshold. At face value, a demand for agency would then encourage participants to cast their votes earlier, to ensure that they have an impact on outcomes.

The argument above is, however, incomplete. Under majority rule, a demand for agency ought to only affect the speed with which the second vote is cast. After the first participant has cast her vote, others may feel pressured to hasten their decisions in order to take part in the process and attain agency. Figure 3, however, illustrates clearly that the posteriors at which *first* voters make decisions are first order stochastically dominated by those corresponding to first voters under unanimity, or in the simulated groups based on our individual treatments. In other words *even first* voters also hasten their decisions under majority rule. How can we reconcile these observations with the simple demand for agency narrative?

It is useful to understand whether the second voter's hastiness is affected by how quickly the first voter votes. In order to evaluate this relationship, we need to attempt to purge artificial connections that are due, for instance, by voters being affected by features of the history of their shared sample paths. Furthermore, as the posterior corresponding to the first vote increases, one may expect the second vote to follow suit more quickly as information more strongly supports the dominant alternative. In order to purge these spurious relationships, we compare how voters vote under majority (and unanimity) with how voters vote in simulated groups. Any connection in simulated groups must be due to the process because the "first" voter is fictitious.

Table 4 displays the results of a regression in which, within each treatment, within each group, within each round, we calculate the difference between the posterior at which the second vote was

[16] See for instance Fehr et al. (2013), Bartling et al. (2014), and Pikulina and Tergiman (2020).

cast, and the posterior at which the first vote was cast.

Recall that d_M and d_U are dummy variables equal to 1 if the data corresponds to our majority and unanimity treatments, respectively. These variables allow for different intercepts across the treatments. The variable p_1 stands for the posterior associated with the first vote cast in the group. The regressors $p_1 \times d_M$ and $p_1 \times d_U$ correspond to the interactions between p_1 and the corresponding treatment dummies, allowing for different slopes across treatments. As in previous regressions, *Last 15* is a variable that equals 1 if the data comes from the last 15 rounds. The variables *Last 15*$\times d_M$ and *Last 15*$\times d_U$ capture the different impacts that learning has on the gap between the first and the second vote across treatments. The same goes for *Slow*$\times d_M$ and *Slow*$\times d_U$—these capture the different impact that a slow process has on the gap between the first and second vote across treatments. To calculate the difference between the posteriors with which votes are cast, we rely on choices across different individuals. Thus, we cluster errors on the process level. Although there is a natural sequencing of votes in our majority and unanimity treatments, no such sequencing exists for the individual treatment. We once more rely on simulating the first, second, and third voters from the individual treatment based on the procedure described in Section 6.[17]

As can be seen from this table, there is no statistically significant difference between either the intercepts or the slopes of the unanimity treatment and the simulated individual treatment. However, d_M and $p_1 \times d_M$ are both statistically significant at the 1% significance level, indicating a different slope and intercept for the majority treatment. In our majority 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.[18]

[17] Since p_1 can take values between 0.5 and 1, before running the regression, we re-normalize all the values of p_1 by subtracting 0.5. Thus, the intercept corresponds to the *additional* accuracy required by the second voter when the first voter casts a vote with a posterior of 0.5.

[18] In Section 11.3.2, 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 at the 10% significance level.

Table 4: Difference in Posterior: Second vs First Voter

	$(p_2 - p_1)$
Constant	0.225***
	(0.0126)
d_M	-0.147***
	(0.0337)
d_U	-0.0251
	(0.0354)
p_1	-0.712***
	(0.0484)
$p_1 \times d_M$	0.154***
	(0.0511)
$p_1 \times d_U$	0.0203
	(0.0454)
Last 15	0.0226***
	(0.00638)
Last 15$\times d_M$	-0.00996
	(0.00661)
Last 15$\times d_U$	0.00733
	(0.00741)
Slow	-0.0463**
	(0.0174)
Slow$\times d_M$	0.00191
	(0.0118)
Slow$\times d_U$	0.0115
	(0.0109)
N	330960

Standard errors in parentheses

Process-level clustering

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

These observations are consistent with the implications of a demand for agency by those who vote second in our majority treatment. Interestingly, given this relation, the first voter can benefit from artificially speeding up the process in order to induce a stopping posterior that is more in line with her preferences. Indeed, this may explain why we observe that the distribution of second voters under majority rule is very similar to the distribution of first voters in simulated groups. The first voter may benefit from manipulating the second voter's response and induce her to stop the group process exactly where the first voter would have chosen had she been by herself.

Certainly, other considerations could explain first voters' hastiness in our majority treatment. For instance, diffusion of responsibility, combined with a demand for agency, could induce an attempt to vote first: one maintains a say in the outcome, but does not conclusively determine it. Such preferences would introduce a race component to participants' strategic interaction and could generate results consistent with those we observe. Nonetheless, they would also suggest an advantage to choosing early in our unanimity treatments, which we do not observe.

7 Static Information Collection

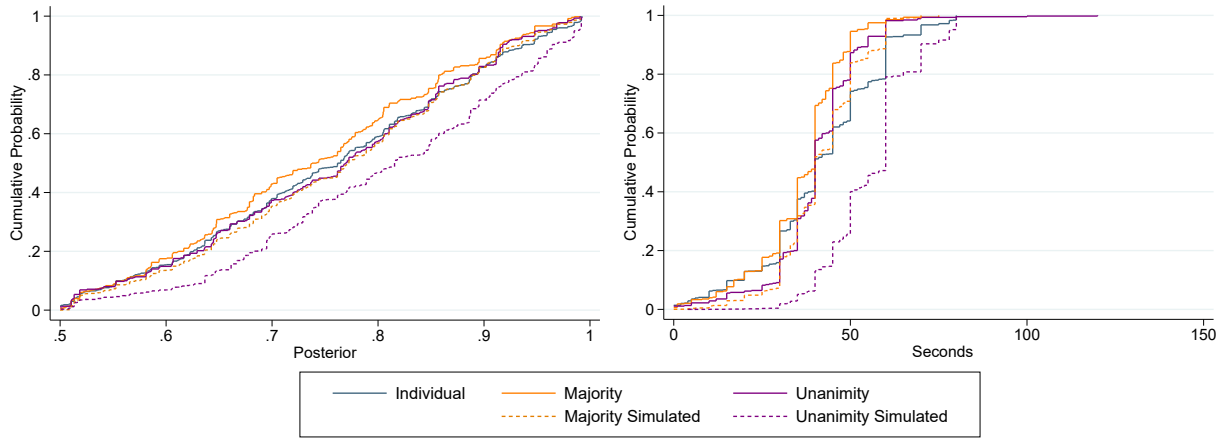
7.1 Group Level Distributions

In the dynamic treatments, our focus was on the posterior probabilities associated with votes. In contrast, in the static treatments, because participants choose the time window for information acquisition, our focus shifts to the time chosen for information collection.

Figure 4 presents the cumulative distribution functions of chosen times across our static treatments. Although in the dynamic case we saw that the distributions were naturally ordered by first

order stochastic dominance, the ordering is less clear. In fact, the distributions of the chosen times appear remarkably similar across the three treatments.

Figure 4: Static Treatment CDFs



The similarity between the distributions across voting rules must be interpreted with caution. The figures show that, in each treatment we observe a lot of heterogeneity in the times chosen by different individuals and groups. Therefore, it must be that individuals behave differently in unanimity from majority or when choosing by themselves. Indeed, absent any explicit "group effect," the mechanical effect presented by groups choosing by majority, selecting the median of three suggested wait times, or unanimity, selecting the maximum of suggested times, would imply differences in distributions. Formally, distributions corresponding to our majority or unanimity treatments should correspond to those of the order statics of the distribution corresponding to our individual treatment. Following the procedure described in Section 6, we simulate the majority and unanimity decision based on our individual-treatment data. We superimpose these resulting cumulative distributions in Figure 4.

Under unanimity, it is the "most patient" group member who governs the group's decisions. It is then perhaps unsurprising that the distribution of wait times derived from the simulated unanimity substantially differs from that corresponding to individual decisions or from the distribution we observe in our unanimity treatment. Indeed, the two-sided Kolmogorov-Smirnov test, presented in Table 5, fails to reject the hypothesis that the distributions associated with the simulated and observed unanimity decisions are different. This test also fails to reject the hypothesis that the majority-simulated distribution differs from the observed majority distribution.

Table 5: Static Treatment Two-Sided Kolmogorov-Smirnov Test

	Majority Treatment			Unanimity Treatment		
	Max Distance	p-value	Corrected	Max Distance	p-value	Corrected
Observed above Simulated	0.2069	0.000		0.5278	0.000	
Simulated above Observed	0.0063	0.963		0.0044	0.982	
Combined K-S	0.2069	0.000	0.000	0.5278	0.000	0.000

Thus, in the static treatments, there seems to be a group effect that goes beyond the purely mechanical one, under both the majority and unanimity voting rules. In particular, both types of collective procedures lead to *hastier* decisions than those generated by our simulated groups.

The regressions reported in Table 6 echo some of the observations above and illustrate the impact of experience in our static treatments. The dummy variable d_M equals 1 if the data comes from the static median treatment, and equals 0 otherwise. The dummy variable d_U takes the value of 1 if the data comes from the static unanimity treatment, and the value of 0 otherwise. The dummy variables *Last 15 I*, *Last 15 M*, and *Last 15 U* equal 1 if the data comes from the last 15 rounds and from the individual, majority, or unanimity treatment, respectively. Each column represents a separate regression, while standard errors are clustered at the individual-level.[19] Versions with no clustering, process level clustering, as well as additional specifications in which we control for the elicited measures of benevolence and risk preferences, are presented in Section 11.3.3 in the Appendix.

Table 6: Static Treatment Group Level Regression

	Seconds Waited		
	All Rounds		Last 15 Rounds
Constant	41.69***	42.92***	40.45***
	(2.309)	(2.244)	(2.716)
d_M	-5.436**	-4.902*	-5.970*
	(2.665)	(2.488)	(3.295)
d_U	-1.222	0.233	-2.676
	(2.582)	(2.475)	(3.311)
Last 15 I		-2.473	
		(1.861)	
Last 15 M		-3.542**	
		(1.482)	
Last 15 U		-5.382***	
		(2.001)	
N	1860	1860	930

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

[19] In the majority and unanimity treatment, clustering is based on the pivotal voter.

In these regressions we see no statistically significant difference between the time waited in the individual and unanimity treatments.[20] In contrast, on average, subjects in the majority treatment chose to wait 5.47 seconds less than the individual treatment, a difference that is statistically significant at the 0.05% significance level. The first regression also make clear that all three treatments yield excessive information collection. Recall from Section 3 that the optimal waiting time is 29.58 seconds. For lower values, the gain in precision outweighs the costs per period, whereas, for values higher than 29.58, the gain in precision is lower than the flow costs participants pay. The deviation from the theoretically-optimal level is quite substantial. In our individual treatment, participants wait 41.69 seconds on average; in our majority treatment, they wait 36.25; and in our unanimity treatment, they wait 39.11 seconds.

Regression results reported in the second column of Table 6 suggest that in all three treatments, experience leads to a reduction in the average chosen time to wait. Since information collection was excessive to begin with, this is a move toward the optimal choice. However, this difference is not statistically significant for the individual treatment when we cluster at the individual level.[21] The average time for the individual, majority, and unanimity treatments drops from 42.92, 38.02, and 43.15, to 40.45, 34.48, and 37.77 respectively. Note that learning is more pronounced in group settings, probably because subjects can observe others choosing lower values, and they decide to experiment with lower values themselves.

Regression results in the third column of Table 6 utilize data only from the last 15 rounds. As can be seen, the estimated parameter values do not change drastically. The gap between the individual treatment and the majority treatment, as well as between the individual and unanimity treatments, grows slightly. Yet, with individual-level clustering, this difference remains statistically insignificant. In other words, although we see some evidence of learning, the extent of learning is limited and, by round 15, participants seem to converge in their behavior.

[20] In Table 11 in the Appendix, where we present the same results with no clustering and with process level clustering, d_U appears statistically significant at the 0.05% or even at the 0.01% in some of the specifications.

[21] In Table 11 in the Appendix, where we present the same results with no clustering and with process level clustering, *Last 15 I* appears statistically significant at the 0.05%.

7.2 Individual Level Distributions

The static and dynamic treatments are similar in that group decisions are sealed by the pivotal vote. There are several differences in how decisions are made, however. First, while in the dynamic treatments participants may change their votes throughout the process (a point we return to below), participants choose only once in the static treatments. Second, while in our dynamic majority treatment we do not observe the choices of those who do not vote before a majority consensus is reached, in the static counterpart, we record all cast votes.

In Figure 5 we present the distribution of the shortest, median, and longest chosen times for the majority and unanimity static treatments. We also present the distribution of the analogous distributions from simulated groups based on the procedure described in Section 6.

Figure 5: Static Treatment CDFs by Vote Order

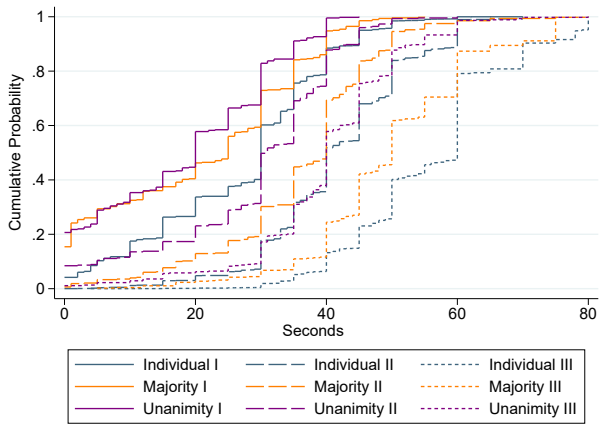


Figure 5 reveals a clear first order stochastic dominance relationship for the shortest, median, and longest times chosen across our treatments. The distributions of all three times corresponding to the unanimity treatment are dominated by those corresponding to the majority treatment, which are dominated by those corresponding to the simulated groups based on the individual treatment. In particular, behavior under both majority and unanimity differs from that in the individual treatments. This echoes our conclusion that group effects are present and go beyond the pure mechanical effects driven by heterogeneity in our sample.

8 Performance

In this section, we compare the performance of individuals and groups in our treatments to shed light on the impact of procedures and decision rules on ultimate outcomes, accounting for both decision quality and information costs. We discuss two alternative ways to evaluate performance. The most first criterion is the welfare of the committee that balances accuracy against the cost of acquiring information. A second criterion is to only consider the accuracy of the decision, which, in our setting, is captured by the posterior with which the committee stops information collection. This is a useful criterion when evaluating the impact of information collection on a broader group of individuals. For instance, juries are composed of a small set of individuals, but the accuracy of verdicts is of interest to the justice system and society at large. Similarly, political committees may encompass a handful of representatives who explore policies that ultimately affect the entire population. We first discuss performance from the point of view of the latter criterion that we call accuracy or "decision quality."

In the theoretical benchmark, 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 .81 in dynamic and .72 in static. However, as we have seen, through in our discussion of aggregate behavior in Section 4.1, in static treatments, on average subjects choose excessively long times in all our treatments, and in individual and majority dynamic treatments, on average subjects terminate information collection too early. Therefore, the observed difference in the quality of decisions between dynamic and static treatments is smaller than predicted. In fact, under majority rule, static information collection leads to superior accuracy on average compared to dynamic information collection. This suggests that, as long as information acquisition costs do not play an important role in the design objective, simpler static procedures do not under perform nearly as badly as the theory predicts.

As we now show, this conclusion changes when we evaluate performance according to the first criterion, which incorporates information acquisition costs. However, this comparison is slightly more complex, and we discuss alternative ways to measure performance, moving from raw measures to more sophisticated approaches that attempt to purge some of the inherent noise.

In each round, participants can potentially receive up to 200 points and, for each minute they

wait, they pay 40 points (where 100 points translate to \$1). We re-normalize the potential 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 when the pivotal vote was cast, we calculate the following performance measure:

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

where i represents a treatment, g represents a particular group in a particular round within the treatment, and λ_i represents the computed performance. That is, we calculate the expected payoff for each group and treatment given the posterior and time at which that group stopped information collection.[22] Table 7 reports regression results that link these performance measures with dummy variables for each treatment. The benchmark is the individual static treatment.[23] The first column reports results for our entire data set, while the second restricts attention to the last 15 rounds in our treatments.

Table 7: Performance Regression

	Performance	
	All Data	Last 15 Rounds
Individual D	0.0401***	0.0381***
	(0.00473)	(0.00585)
Majority D	0.0347***	0.0373***
	(0.00569)	(0.00667)
Unanimity D	0.0471***	0.0416***
	(0.00696)	(0.00892)
Majority S	0.00196	0.00478
	(0.00603)	(0.00812)
Unanimity S	0.00508	0.00342
	(0.00540)	(0.00797)
Constant	0.615***	0.616***
	(0.00325)	(0.00399)
N	3840	1920

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

In line with theoretical predictions, all of our dynamic treatments generate significantly higher performance than all of our static treatments. The average performance of the majority and

[22] Experimental payoffs, taking into account the realized states corresponding to each process, yield similar comparisons. We present the expected payoffs in order not to confound conclusions with the inherent randomness of outcomes in our limited number of 30 rounds.

[23] That is, estimated coefficients represent the difference of the mean performance of other treatments from the individual static treatment. The mean performance of the individual static treatment is captured by the *Constant*.

unanimity static treatments is higher than the mean performance of the individual static treatment and the average performances of the individual and unanimity dynamic treatment is higher than the mean performance in the dynamic majority treatment. However, these differences are not statistically significant. Learning does not seem to have important impacts on performance: results from the last 15 rounds resemble those from the entire data set.

The performance measure assessed above necessarily inherits a certain element of randomness imposed by the particular information processes participants face. Take, for example, our static treatments. Ultimately, participants select the time at which information collection stops. The precise posterior that time generates depends on the realized process. This inherent randomness introduces noise in our assessments, which could render comparisons between treatments insignificant. When considering performance, one may then be interested in the *expected* welfare, accounting for the *expected* decision accuracy implied by each choice of stopping time. Similarly, in our dynamic treatments, it is natural to consider the *expected* time induced by any choice of posterior and assess performance accordingly. We now discuss how to use the theory to obtain measures that are less subject to this noise.

For the static case, from the analysis in Section 3.1, we know that $t^* = 29.58$ and that the expected posterior given any t is $\mathbb{E}[p|t] = \frac{1}{2}\left(\text{erf}\left(\frac{\sqrt{\mu t}}{2}\right) + 1\right)$. The expected performance under the optimal stopping time is then:

$$\lambda_{static}^{theory} = \mathbb{E}[p|t^*] - ct^* = \frac{1}{2}\left(\text{erf}\left(\frac{\sqrt{\mu t^*}}{2}\right) + 1\right) - ct^* = 0.623$$

For the dynamic case, from Section 3.2, we know that the optimal threshold is $p^* = 0.81$. The expected stopping time, given any fixed threshold p, is $\mathbb{E}[t|p] = \frac{(2p-1)\log\left(\frac{p}{1-p}\right)}{\mu}$.[24] The expected performance when choosing the optimal posterior threshold is then:

$$\lambda_{dynamic}^{theory} = p^* - c\mathbb{E}[t|p^*] = p^* - \mathbb{E}[t|p^*] = p^* - c\frac{(2p^* - 1)\log\left(\frac{p^*}{1-p^*}\right)}{\mu} = 0.680$$

[24] For simplicity, we effectively assume a stationary threshold here.

These expected performance values offer upper bounds on expected performance.

Consider now an immediate decision, corresponding to a choice of posterior 0.5 in the dynamic treatments, or a choice of 0 waiting time in the static treatments. Immediate decisions generate the correct guess 50% of the time and come at no information costs. This benchmark constitutes a plausible lower bound on performance and results in an expected payoff of $\underline{\lambda} = 0.5$.[25] We now construct a new performance measure as follows:

$$\tilde{\lambda}_i = \frac{\lambda_i - \underline{\lambda}}{\lambda_i^{Theory} - \underline{\lambda}} \qquad i \in \{Static,\ Dynamic\}$$

This new measure captures the relative performance of each treatment between the "worst" and the theoretically best performance. A score of 0 would indicate that, on average, the treatment performs no better than an immediate decision that incorporates no information. A score of 1 would indicate that, on average, the treatment exhibits optimal performance.

With this measure, we see significant differences between our various dynamic treatments and our various static treatments. Relative performance in the dynamic unanimity treatment is statistically higher than that observed in our individual and majority dynamic treatments ($p < 0.05$); Relative performance in the static majority treatment is statistically higher than that observed in our individual and unanimity static treatments ($p < 0.05$).[26]

[25] It is certainly possible to achieve lower performance. For example, an excessively long wait can yield negative expected payoffs. However, we do not observe such behavior in the lab.

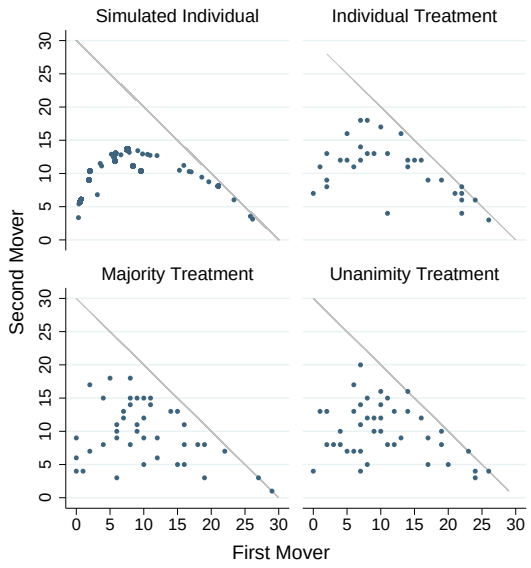
[26] Specifically, the assessed performance in our dynamic individual, majority, and unanimity treatments are 0.862, 0.832, and 0.901, respectively. The assessed performance in our static individual, majority, and unanimity treatments are 0.935, 0.951, and 0.976, respectively. Nearly identical values are observed in the last 15 rounds.

9 Additional Features of Individual Behavior

9.1 Individual Voting Order

Figure 6 describes the voting order of our subjects across all rounds.

Figure 6: Dynamic Treatments Voting Order

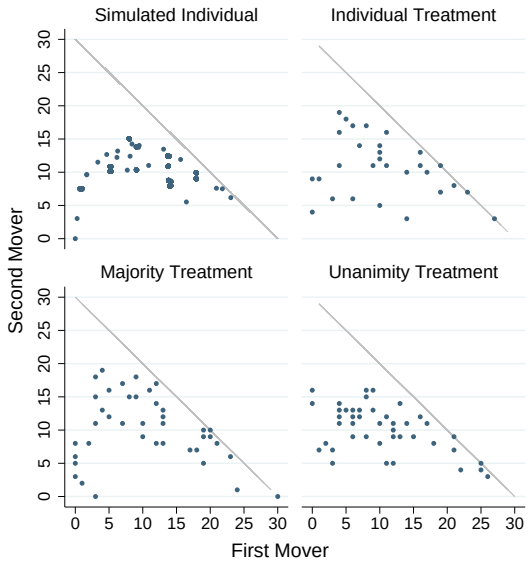


The horizontal axis represents the number of rounds the participant is the first voter, whereas the vertical axis represents the number of rounds the participant is the second voter. If a participant was never the third voter, or in the majority case, a round never ended without them voting, then they would lie on the gray line. The further away they are from the gray line, the higher the number of rounds they were the third voter, or in the majority treatment, the higher the number of times they did not get to vote. We observe a lot of heterogeneity in voting order, but there is some degree of persistence in behavior.

For the individual treatment, of course there are no groups and hence no first, second or third voters. However we present the distribution of vote orders by randomly grouping individuals who voted in isolation. The "Individual Treatment" presents one such random grouping, thus the number of observations is on the order of the number of observations in Majority and Unanimity treatments. On the other hand the "Simulated Individual" represents these shares from 30,000 simulated groups.

Figure 7 performs a similar exercise for our static treatments and we see similar patterns.

Figure 7: Static Treatments Voting Order



9.2 Multiple Voting

The participants initial position is W, which stands for wait. Throughout the game they can choose to move to A or B, and depending on the treatment, if a pivotal vote has not been cast yet, they can move back to W, or jump fro A to B directly, or vice-versa, if they so choose. These options are of course not available in the individual treatment. In the individual treatment once the participant moves from W to A or B the game immediately ends. The interesting cases then are unanimity and majority treatments.

Figure 8 represents the number of times a participant cast at least one, two, three, and so on, votes. In the unanimity treatment, a decision can not be made without each participant casting at least one vote, which is why the number of times participants cast at least one votes is $48 \times 30 = 1440$.

Figure 8: Voting Times by Treatment

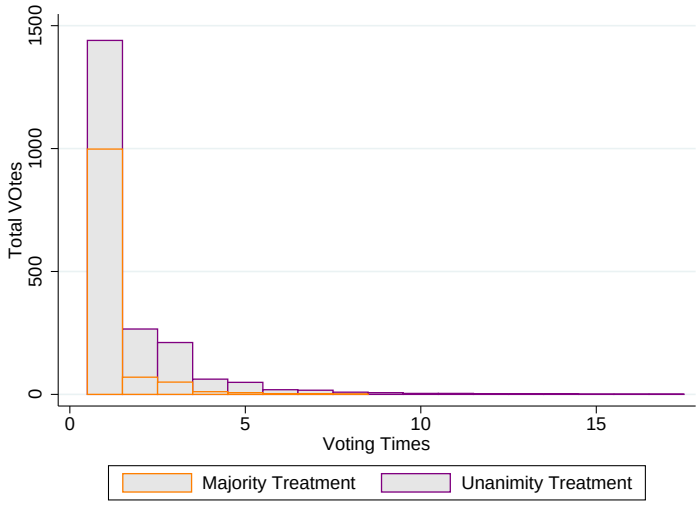
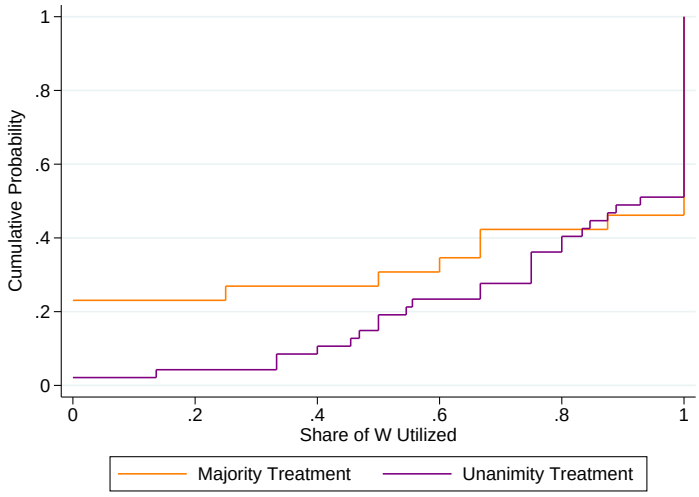


Figure 9 represents the share of times a participant who previously cast a vote on A or B, and votes at least one more time, follows this vote with a vote on W. As can be seen, in the majority treatment, there is a mass point on 0, implying that some participants never utilize W and thus jump from A to B, or vice-versa, directly. In both the majority and unanimity treatments there is a mass-point on 1, implying that these participants always follow a vote on A or B with a vote on W.

Figure 9: W Utilization by Treatment



10 Conclusions

This paper documents experimentally how individuals and groups collect information under various institutional constraints governing both the timing of decisions and the procedures by which opinions are aggregated. There are several important insights. Overall, our participants come remarkably close to the theoretically optimal information-collection policy. However, groups deciding via majority dynamically are far hastier than both individuals and groups using unanimity. Furthermore, when individuals or groups collect information dynamically, they become more lenient over time—they make less accurate decisions as time goes by. As theory suggests, dynamic information collection generates greater utilitarian welfare than static information collection for deciding bodies. Nonetheless, excessive information collection in static treatments yields more accurate decisions in those settings than in dynamic ones.

Taken together, our study provides some guidance for the design of decision protocols. When the deciding body impacts a large segment of the population, as is the case for juries, political committees, etc., decision accuracies may be of utmost importance and static information collection may be desirable. In other cases, or when information collection inherently takes place over time, decisions made by individuals or by groups with more stringent voting rules may yield superior outcomes.

11 Appendix

11.1 Beyond Risk-Neutrality

11.1.1 Static Version

Let:

$$p(t) := \frac{1}{2}\left(\text{erf}\left(\frac{\sqrt{\mu t}}{2}\right) + 1\right)$$

Then the problem for the static case becomes:

$$\max_t \; p(t)u(x - ct) + (1 - p(t))u(-ct)$$

Where x represents the reward, c represents the cost, t represents time the participant decides to wait, and $u(\cdot)$ is the utility function of the agent. The first order condition leads to:

$$\frac{u(x - ct) - u(-ct)}{p(t)u'(x - ct) + (1 - p(t))u'(-ct)}p'(t) = c \tag{1}$$

Where:

$$p'(t) = \frac{\mu e^{-\frac{1}{4}(\mu t)}}{4\sqrt{\pi}\sqrt{\mu t}}$$

Given that:

$$u(x - ct) = u(-ct) + \int_{-ct}^{x-ct} u'(s)ds$$

The above can be written as:

$$\frac{\int_{-ct}^{x-ct} u'(s)ds}{p(t)u'(x - ct) + (1 - p(t))u'(-ct)}p'(t) = c$$

Which reduces to $x\,p'(t) = c$ in the risk neutral case. The part multiplying $p'(t)$ is not always lower or greater than one for any x, c, t. Thus, it is not clear whether risk averse agent chooses to wait more or less than a risk neutral agent.

Example I That a risk averse agent does not necessarily wait more or less than a risk neutral agent can be seen in the following example:

$$u(z) = \gamma\left(1 - e^{-z}\right) + (1-\gamma)z$$

For $\gamma = 0$ we are back to the risk neutral case, and for values of $\gamma > 0$ the utility function has curvature. We employ this utility function as it is well defined even for negative values. We then have:

$$\begin{aligned}\mathbb{E}\left[u(z)\right] &= \sum_i u(z_i)p_i \\ &= p(t)\left(\gamma\left(1 - e^{-(x-ct)}\right) + (1-\gamma)(x-ct)\right) + (1-p(t))\left(\gamma\left(1 - e^{-(-ct)}\right) + (1-\gamma)(-ct)\right)\end{aligned}$$

Plugging in $p(t)$ which was defined earlier, the above expression becomes:

$$\mathbb{E}\left[u(z)\right] = \frac{1}{2}\left(\operatorname{erf}\left(\frac{\sqrt{\mu t}}{2}\right)\left(\gamma\left(e^{x}-1\right)e^{ct-x} - \gamma x + x\right) + 2c(\gamma-1)t - \gamma\left(e^{x}+1\right)e^{ct-x} + 2\gamma - \gamma x + x\right)$$

For any value of $\gamma > 0$ there is no closed form solution. Whereas, for $\gamma = 0$, the optimal solution from part Section 3 holds. With $x = 1$, $\mu = 1.4$, and $c = 0.2$, with no risk aversion $\gamma = 0$, optimal waiting time will be $t^* = 0.49$, corresponding to 29.6 seconds. On the other hand, if $\gamma = 1$, we numerically find that the expected utility is maximized when $t^* = 0.64$, corresponding to 38.4 seconds. Furthermore, for the given parameter values x, μ, and c, as γ increases from 0 to 1, the optimal waiting time monotonically increases.

Consider now the same problem with reward $x = 3$, cost $c = 0.7$, and unchanged drift $\mu = 1.4$. Once more, from the closed form solution in the risk-neutral case, $\gamma = 0$, we find that the optimal waiting time is $t^* = 0.38$ while numerically we find that the optimal waiting time when $\gamma = 1$ reduces to 0.33. Furthermore, for the given parameter values, as γ increases from 0 to 1, the optimal waiting time now monotonically decreases.

Example II Consider yet another example with CRRA utility function. Let:

$$u(z) = \frac{1}{1-\theta} z^{1-\theta} \qquad \theta > 0$$

We put no further restrictions on θ. For values of z lower than 0, the above function will not be well defined. In particular, the expected utility will be:

$$\mathbb{E}\left[u(z)\right] = p(t)\left(\frac{1}{1-\theta}(x-ct)^{1-\theta}\right) + (1-p(t))\left(\frac{1}{1-\theta}(-ct)^{1-\theta}\right)$$

To avoid ending up with a negative value, we can add an additional payoff of y in both states, this can be thought of as the show-up-fee in the experiment. As long as $t \leq \frac{y}{c}$ both states will have non-negative payoffs. However, adding y is not without loss of generality. Shifting both states by y changes the agents risk attitude. Nonetheless, we proceed with the analysis in this fashion. The expected payoff is then:

$$\mathbb{E}\left[u(z)\right] = p(t)\left(\frac{1}{1-\theta}(x+y-ct)^{1-\theta}\right) + (1-p(t))\left(\frac{1}{1-\theta}(y-ct)^{1-\theta}\right)$$

The first order condition leads to:

$$\frac{1}{4}\left(-2c\left(\text{erf}\left(\frac{\sqrt{\mu t}}{2}\right)+1\right)(-ct+x+y)^{-\theta} - 2c\text{erfc}\left(\frac{\sqrt{\mu t}}{2}\right)(y-ct)^{-\theta}\right)$$
$$+\frac{1}{4}\left(+\frac{\mu e^{-\frac{1}{4}(\mu t)}(-ct+x+y)^{1-\theta}}{\sqrt{\pi}(1-\theta)\sqrt{\mu t}} + \frac{\mu e^{-\frac{1}{4}(\mu t)}(y-ct)^{1-\theta}}{\sqrt{\pi}(\theta-1)\sqrt{\mu t}}\right) = 0$$

If $\theta = 0$, optimal t is once more in line with Section 3, and thus for $x = 1$, $c = 0.2$ and $\mu = 1.4$ we have $t^* = 0.49$. For any value of $\theta \neq 0$, there is no closed form solution for optimal t. Let $y = 0.3$, allowing for $t \in [0, 1.5]$ without introducing a negative payoff in any state. We numerically find that the optimal waiting time with $\theta = 0.2$ is $t^* = 0.51$, while with $\theta = 0.8$, thus, with more risk aversion, the optimal waiting time is $t^* = 0.46$. Hence, even with a $CRRA$ utility function there seems to be no monotonicity with respect to the effect that risk aversion has on the optimal waiting time.

To summarize, for non risk-neutral agents, the new optimally condition is defined by equation 1. Since the part multiplying $p'(t)$ for different parameter values may be higher or lower than one,

it is not clear whether risk averse agents wait more or less than risk neutral agents. The particular examples show that for risk averse agents both longer and shorter optimal choices are possible. To build intuition of why this behavior occurs, note that to reduce uncertainty the agents have to wait longer. However, waiting longer shifts both payoffs downward, both $-ct$ and $x - ct$ decrease as t increases. A higher risk aversion might make it beneficial to decrease payoffs in both states for the sake of more certainty. However, higher risk aversion, from the additional curvature, makes ending up with the lower $-ct$ state more costly, pushing towards the other side of the trade-off. Hence, depending on the particular utility function and the parameter values, a more risk averse agent would either find it optimal to decrease uncertainty, choose to wait more, or choose to wait less so that the bad outcome is less painful.

11.1.2 Dynamic Version

Consider the individual dynamic case. Lets just assume that a threshold equilibrium is being played, where the threshold is equal to $\tilde{p}$. This threshold then gives rise to a distribution of end times, $f(t|\tilde{p})$ (for which only a Fourier series representation can be constructed). For any $\hat{t}$ in which the game ends, the agent then receives the following lottery:

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

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

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

So then, by choosing a larger $\tilde{p}$ the agent minimizes the uncertainty in the lottery she receives, however this increases the uncertainty regarding the ending time. It is not clear what a risk averse agent would prefer, or that such a preference would be stable throughout different parameter values.

11.2 Pulling the Trigger

Figure 10: Pulling the Trigger: Majority Treatments

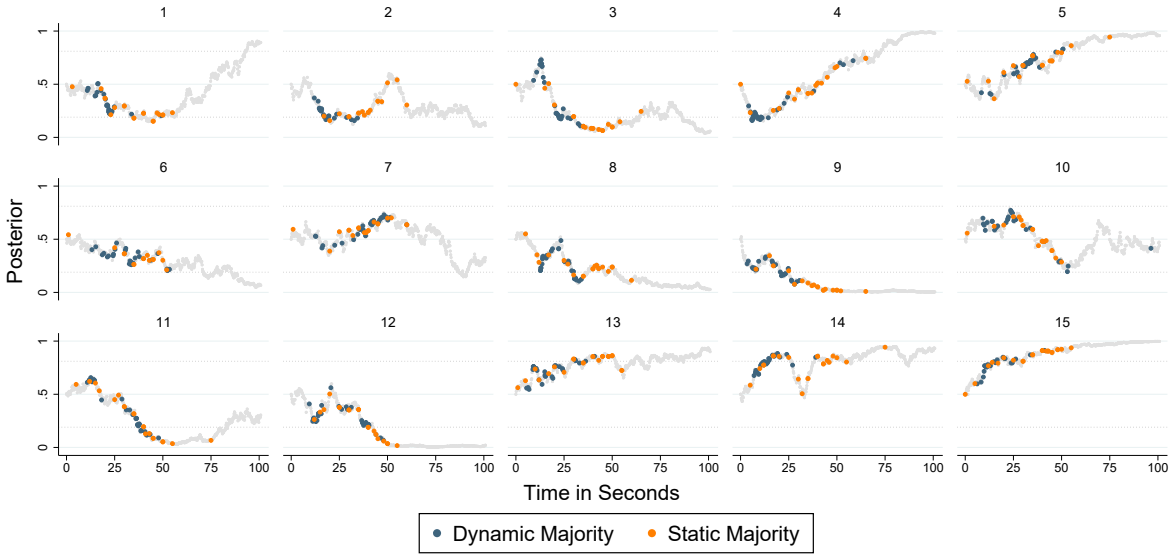
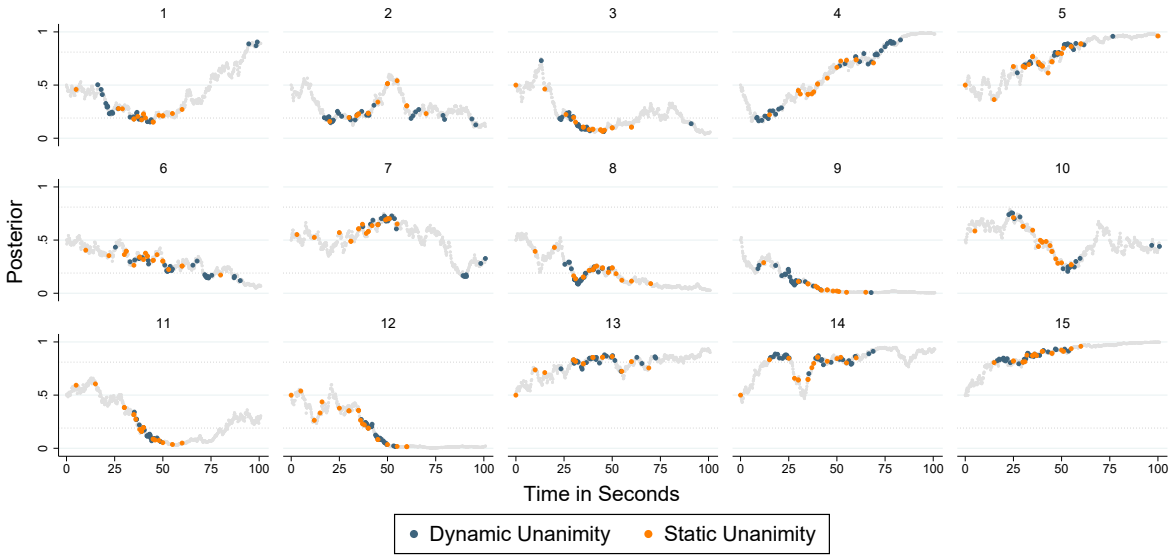


Figure 11: Pulling the Trigger: Unanimity Treatments



11.3 Additional Analysis and Alternative Specifications

11.3.1 Dynamic Treatment Regressions

In this section we analyze alternative specifications of the dynamic treatment regression presented in Table 3. In the first regression in Table 8, first column, continues to cluster standard errors on the individual level while introducing two new variables as regressors, *Tokens Sent* and *Tokens Not Invested*. As mentioned in Section 2, at the end of each session, participants completed two risk-elicitation tasks as in Gneezy and Potters (1997). Namely, participants had 200 tokens to invest in a safe or risky asset. Tokens that were not invested were kept in the safe asset. The variable *Tokens Not Invested* which can have a value between 0 and 200, represents the amount participants did not invest on the risky asset.[27] Thus, roughly speaking, the higher this value is, the more risk averse the participant seems. At the end of each session participants also played a dictator game, in which they were given 200 tokens and decided how much to keep for themselves, and how much to give to another participant with whom they have been randomly paired. The variable *Tokens Sent* represents the amount of tokens the participant gave to the matched paired partner.[28] Since we elicit each measure twice, we run an instrumental variable regression, using the first elicitation as an instrument for the second. Doing so accounts for the fact that these are noisy elicitations.

Table 8: Dynamic Treatment Alternative Specifications

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

Standard errors in parentheses

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

Neither *Tokens Sent*, nor *Tokens Not Invested* appear statistically significant. On the other

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

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

hand, the sign and magnitude of all other estimated parameters remains roughly unchanged.

The following three regressions/columns in Table 8 represents regressions akin to the ones found in Table 3 with and without *Tokens Sent* and *Tokens Not Invested*. The difference here is that standard errors have not been clustered. Whereas the last three regressions/columns in Table 8 represent the same analysis with standard errors clustered at the process level. As explained in Section 2, we draw a total of 15 Wiener processes with a drift, each utilized twice. It is at this process level that we cluster in the last four columns. The statistical significance with respect to the initial parameters remains unchanged. Whereas *Tokens Sent* now appears statistically significant. However, considering that about 60% of the participants send a value of 0 tokens, and more than 80% send less than 50 tokens, given the estimated parameter value, this variable does not seem to have the potential to explain a substantial portion of the variation in the posterior. The same argument can be made with regards to *Tokens Not Invested*, as that value can range from 0 to 200.

Table 9 presents regression results identical to the ones presented in Table 3, with process-level clustering or no clustering. Notice that the fixed effects regression can not be presented with process level clustering as the panels are not nested within clusters.

Table 9: Decreasing Thresholds Alternative Clustering

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

Standard errors in parentheses

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

The only noticeable difference from Table 3 seems to be the weakening, or loss of the statistical significance of *Slow M* under process level clustering.

11.3.2 Difference in Posterior: Third vs Second Voter

Table 10 presents a regression similar to the regression presented in Table 4. The dependent variable here is the difference between the posterior of the third and second vote. Furthermore, since in the majority treatment only two votes are required for a decision to be made, this regression utilizes data only from the unanimity treatment as well as the simulated individual treatment.[29]

Table 10: Difference in Posterior: Third vs Second Voter

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

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

As can be seen, the there is no statistically significant difference of the intercept between the simulated individual treatment and the unanimity treatment. Whereas, the $p_2 \times d_U$ is statistically significant at the 10% significance level. Indicating that the unanimity treatment may have a slightly flatter slope than the simulated individual treatment. However, since it's intercept is also lower, the difference between the two remains rather small.

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

11.3.3 Static Treatment Regressions

In this section we analyze alternative specifications of the static treatment group level regression presented in Table 6. The first three regressions/columns, and the last three regressions/columns of Table 11 have the same specification as the three regressions/columns in Table 6. The difference is the level on which we cluster the standard errors. In Table 6 we presented individual level clustered standard errors, where the individual was the pivotal vote caster. In Table 11, in the first three regressions/columns we present the results with no clustering, whereas in the last three regressions/columns we present the results with process level clustering. As explained in Section 2, we draw a total of 15 Wiener processes with a drift, each utilized twice. It is at this process level that we cluster in the last three columns. Compared to the individual level clustering, a in either case presented here, almost all parameter values show an increase in statistical significance.

Table 11: Static Treatment Group Level Regression Alternative Clustering

	Seconds Waited					
	No Clustering			Process Level Clustering		
	All Rounds		Last 15 Rounds	All Rounds		Last 15 Rounds
Constant	41.69***	42.92***	40.45***	41.69***	42.92***	40.45***
	(0.492)	(0.691)	(0.743)	(0.576)	(0.970)	(0.430)
d_M	-5.436***	-4.902***	-5.970***	-5.436***	-4.902***	-5.970***
	(0.843)	(1.184)	(1.273)	(0.470)	(0.730)	(0.462)
d_U	-1.222	0.233	-2.676**	-1.222**	0.233	-2.676***
	(0.861)	(1.210)	(1.300)	(0.432)	(0.601)	(0.865)
Last 15 I		-2.473**			-2.473**	
		(0.977)			(0.962)	
Last 15 M		-3.542***			-3.542***	
		(1.360)			(0.670)	
Last 15 U		-5.382***			-5.382***	
		(1.404)			(1.060)	
N	1860	1860	930	1860	1860	930

Standard errors in parentheses

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

The regressions presented in Table 12 introduces two new variables as regressors, *Tokens Sent* and *Tokens Not Invested*. What these variables represent has been described in Section 11.3. Since we elicit each measure twice, we run am instrumental variable regression, using the first elicitation as an instrument for the second. Doing so accounts for the fact that these are noisy elicitations.

Table 12: Static Treatment Group Level Regression with Benevolence and Risk Preferences

	Seconds Waited					
	All Rounds			Last 15 Rounds		
	No Clustering	Individual Clustering	Process Clustering	No Clustering	Individual Clustering	Process Clustering
Constant	42.67***	42.67***	42.67***	43.95***	43.95***	43.95***
	(1.465)	(7.633)	(0.903)	(2.098)	(8.338)	(0.846)
d_M	-4.555***	-4.555	-4.555***	-5.717***	-5.717	-5.717***
	(1.286)	(4.483)	(0.790)	(1.498)	(5.090)	(0.457)
d_U	0.687	0.687	0.687	-1.927	-1.927	-1.927*
	(1.368)	(4.910)	(0.499)	(1.641)	(5.495)	(1.017)
Last 15 I	-2.473**	-2.473	-2.473***			
	(0.972)	(1.851)	(0.928)			
Last 15 M	-3.542***	-3.542**	-3.542***			
	(1.353)	(1.480)	(0.646)			
Last 15 U	-5.382***	-5.382***	-5.382***			
	(1.397)	(1.966)	(1.023)			
Tokens Sent	-0.0323	-0.0323	-0.0323	-0.106	-0.106	-0.106***
	(0.0476)	(0.210)	(0.0248)	(0.0733)	(0.239)	(0.0314)
Tokens Not Invested	0.00424	0.00424	0.00424	-0.0256	-0.0256	-0.0256***
	(0.0123)	(0.0659)	(0.00518)	(0.0188)	(0.0706)	(0.00760)
N	1860	1860	1860	930	930	930

Standard errors in parentheses

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

The first three regressions/columns presented in Table 12 utilize the whole data, whereas the last three regressions/columns utilize the data from the last 15 rounds only. Not clustered, individual level clustered, and process level clustered standard errors are presented in each case in separate columns. As can be seen, *Tokens Sent* and *Tokens Not Invested* appear statistically insignificant in all but the last regression/column, in which, data from the last 15 rounds has been utilized and errors are clustered at the process level. The estimated parameter value for *Tokens Sent* is -0.106, while for *Tokens Not Invested* the estimated parameter value is -0.0256. Thus, according to this specification, more "benevolent" participants tend to choose lower waiting times, as do more "risk averse" participants. Initially, intuitively, it might seem odd that more "risk averse" participants choose lower waiting times. However, as we discuss in Section 11.1, it is not straightforward how risk aversion impacts the optimal waiting time. In some specifications, more risk aversion leads to shorter optimal waiting times, while in other specifications, or for alternative parameter values, longer waiting times may be optimal, relative to risk-neutral agents.

11.4 Learning

11.4.1 Dynamic Treatment Learning

We begin our study of how decisions change throughout the experiment by examining whether there is a trend in the posterior with which participants cast their individual votes. In Table 13 we regress

the posterior with which individuals cast their individual vote on *Round*, which stands for the game round, *Slow*, which identifies the process occurring during the round as a slow or a quick process, and an interaction between *Round* and *Slow*, allowing for a different learning trend depending on the process. [30] We run a 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 with different magnitudes. 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 $Correct_{t-1}$ which is equal to 1 if the last period individual decision, or group decision in the majority and unanimity treatment, was correct, and equal to 0 if the decision was incorrect. And finally we control for $Difference_{t-1}$, which is equal to the difference of the participants last period choice from the mean of other participant's choices in the last period. This can be calculated for the majority and unanimity cases only.

Table 13: 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
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

From the estimated coefficients on $Correct_{t-1}$ throughout all the specification, we see that on average participants cast their individual votes with a lower posterior in round t if their/their group's guess in round $t-1$ was correct, compared to the posterior with which they typically cast their votes when their/their group's guess guess was wrong in round $t-1$. In contrast, the coefficients on $Difference_{t-1}$ is never statistically significant, implying that participants are not

[30] Additionally what motivates us to allow for a different intercept and slope for quick and slow processes is to avoid falsely identifying a trend where there is none. We showed earlier that participants tend to vote with a lower posterior when faced with a slow process compared to a quick processes. If in earlier rounds there are more quick processes, whereas in later rounds there are more slow processes, had we not allowed for a separate intercept and slope, it would appear that the average posterior has gone down thought the rounds, or vice versa.

greatly affected by the decisions of other participants in the previous round.

The main focus in this analysis is to compare the magnitude and statistical significance of *Round* and *Round* × *Slow* in the first 15 and last 15 rounds. By doing so we aim to understand whether participants had enough rounds to learn and adjust their strategies. In our view, the best setting to evaluate this, is the individual treatment, as each choice made is the final implemented decision, whereas in the other two treatments, the individual decision might not necessarily be the pivotal one. However, we present the results for each treatment. The regressions in column two and three reveal that both the magnitude and statistical significance of *Round* and *Round* × *Slow* drops in the last 15 rounds compared to the first 15 rounds in the individual treatment. Similar comparisons for the majority treatment, column five and six, and the unanimity treatment, column eight and nine, reveal a decrease in statistical significance and magnitude. Even in the cases where statistical significance persists, the magnitude is much lower in the last 15 round. For example, in the majority treatment, the slope is highest in the first 15 rounds when the process is slow $Round + (Round \times Slow) = 0.00177 + 0.0110 = 0.01277$, which drops to 0.00455 in the last 15 rounds. In the unanimity treatment the most significant slope in the first 15 rounds is also for slow processes $Round + (Round \times Slow) = 0.00267 + 0.00904 = 0.01171$, which in the last 15 rounds drops to 0.00114. The finding that the magnitude of learning tends to be much 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 in our experiment there would not be much value to allowing subjects to have additional rounds.

11.4.2 Static Treatment Learning

We now perform a similar analysis for the static treatment where we examine whether there is a trend in the decisions participants make throughout the rounds. The specification of the regressions presented in Table 14 is as described in Section 11.4.1. However the dependent variable, the participant's main choice, is now time waited (ALESSANDRO: IS THIS A GOOD NAME?), instead of the posterior. Furthermore, in the static case participants cannot react differently to slow and quick processes, because the process evolves only after decisions have been made. Thus, we do not include *Slow* and *Round* × *Slow* in the regression below.

Table 14: 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

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

The estimated coefficients on $Correct_{t-1}$ and $Difference_{t-1}$ are typically insignificant, implying that 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 member's last period decision. Once more, the magnitude and the statistical significance of *Round* seems to be greatly reduced in the last 15 rounds compared to the initial 15 rounds. This leads us to believe that participants had sufficient rounds to learn and adjust their decisions.

11.5 Alternative Performance Measures

To compute the performance measure in Section 8 we utilized the posterior p and time t when the group decision was made. In this section we present two alternative performance measures. To construct the firs new performance measure we will utilize the actual realization, that is, whether or not the group guessed correctly, instead of the posterior p, which represents the probability with which the group would have guessed correctly. Thus, we compute:

$$\lambda_{i,g}^{realized} = c_{i,g} - 0.2t_{i,g}$$

Where $c_{i,g} \in \{0, 1\}$ represents whether or not in a particular round, a particular group g in treatment i guessed correctly. This performance measure is influenced by "luck" much more than the performance measure presented in Section 8. For example, a group which cast the pivotal vote with a posterior of 0.90 within say 30 seconds, would have ended up scoring $0.90 - 0.2 \cdot 0.50 = 0.80$ under the previous measure, whereas with the new measure, if by chance the group's guess turned

out to be incorrect, they would be assigned a score of $0 - 0.2 \cdot 0.5 = -0.10$. While the influence of luck is smaller in the previous performance measure, it still has an impact. For example, consider a group in the static treatment that decides to wait for 30 seconds. The expected posterior for this group is $\mathbb{E}[p|t] = \frac{1}{2}\left(\text{erf}\left(\frac{\sqrt{\mu t}}{2}\right) + 1\right) = \frac{1}{2}\left(\text{erf}\left(\frac{\sqrt{1.4 \cdot 0.5}}{2}\right) + 1\right) = 0.72$. While, by chance, for this group the resulting posterior might turn out to be 0.55 or perhaps the group got lucky and the resulting posterior turned out to be 0.95. Thus, the performance measure calculated in Section 8 is influenced not only by the actions of the group, but by random chance as well.

Motivated by these observations we construct a third and final performance measure. To compute this performance measure, for the static treatment, realizing that the choice variable is simply the time the group waits, instead of utilizing the realized posterior value, we utilize the expected posterior value. Thus, for the static treatment we compute the new performance measure as follows:

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

Similarly for the dynamic treatment, at least in theory, participants choose when to cast their votes based on the posterior they observe. If a participant decides to cast their vote with a posterior of 0.80, sometimes they might get lucky and achieve this posterior within a few seconds, whereas, in other cases they might have to wait a rather long time before observing their desired posterior level. To reduce the impact that luck may have on these performance measures, for the dynamic treatment, for the new measure we will utilize the expected time given a posterior, rather than the actual time.

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

Table 15 presented below, is similar to Table 7, however, it presents the estimated parameter values by utilizing the two new performance measures.

Table 15: Alternative Performance Regression

	Performance			
	All Rounds		Last 15 Rounds	
	$\lambda^{realized}$	$\lambda^{expected}$	$\lambda^{realized}$	$\lambda^{expected}$
Individual D	0.0499***	0.0428***	0.0618***	0.0426***
	(0.0131)	(0.00362)	(0.0182)	(0.00419)
Majority D	0.0423*	0.0390***	0.0415	0.0421***
	(0.0225)	(0.00347)	(0.0293)	(0.00363)
Unanimity D	0.0386**	0.0514***	0.0730***	0.0503***
	(0.0184)	(0.00314)	(0.0251)	(0.00371)
Majority S	0.000513	0.00638**	0.0161	0.00470*
	(0.0205)	(0.00248)	(0.0299)	(0.00283)
Unanimity S	-0.0171	0.00506**	-0.000399	0.00435
	(0.0229)	(0.00243)	(0.0282)	(0.00280)
Constant	0.591***	0.609***	0.586***	0.609***
	(0.0103)	(0.00213)	(0.0144)	(0.00221)
N	3840	3840	1920	1920

Standard errors in parentheses

Individual-level clustering

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

Once more, as in Table 7, the dynamic treatments seem to outperform the static treatments. However, the significance levels are somewhat lower with the $\lambda^{realized}$ performance measure. This is, of course not surprising, as as was stated above, this performance measure is impacted by luck more than any other performance measure, and as such, it will be much noisier than the other two.

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*.

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

Bhattacharya, S., Duffy, J., and Kim, S. (2013). Information acquisition and voting mechanisms: Theory and evidence. *mimeo*.

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.

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. (2018). 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.

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.

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. (2016). Ignorance and bias in collective decisions. *Journal of Economic Behavior & Organization*.

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.

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.

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*.

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. (2017). Experimentation and approval mechanisms. *Unpublished working paper*.

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.

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. (2015). Costly information acquisition in a speculative attack: Theory and experiments. *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.