Essays on social learning and social choice

Thesis by
Wade Hann-Caruthers

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended May 1, 2023

Wade Hann-Caruthers
ORCID: 0000-0002-4273-6249

ABSTRACT

This dissertation contains three essays, two which contribute to the study of social learning (Chapters 1 and 2) and one which contributes to the study of social choice (Chapter 3).

In Chapter 1, I introduce a fully rational model of social learning on networks with endogenous action timing. I show that the structure of the network can play an important role in the aggregation of information. When the social network contains high-degree vertices, agents can be arbitrarily likely to make good choices. In contrast, when the social network is linear, there is a bound on how likely agents are to make good choices which holds regardless of how patient they are. The main contribution of this chapter is the identification of a novel mechanism through which strategic behavior can substantially impede the flow of information through a social network.

In Chapter 2, co-authored with Vadim Martynov and Omer Tamuz, we study the asymptotic rate at which the probability of taking the correct action converges to 1 in the classical sequential learning model with unbounded signals. We provide a characterization of the asymptotic law of motion of the public belief, and we use this characterization to show that convergence occurs more slowly than when agents directly observe private signals, and that the expected time until the last incorrect action can be finite or infinite.

In Chapter 3, co-authored with Laurent Bartholdi, Maya Josyula, Omer Tamuz, and Leeat Yariv, we introduce *equitability* as a less stringent alternative to symmetry for modeling egalitarianism in voting rules. We then use techniques from group theory to show that equitable voting rules can have minimal winning coalitions comprising a vanishing fraction of the population, but they cannot be smaller than the square root of the population size.

PUBLISHED CONTENT AND CONTRIBUTIONS

Laurent Bartholdi, Wade Hann-Caruthers, Maya Josyula, Omer Tamuz, and Leeat Yariv. Equitable voting rules. *Econometrica*, 89(2):563–589, 2021. doi: 10.3982/ECTA17032.

All authors contributed equally at all stages of this project.

Wade Hann-Caruthers, Vadim V Martynov, and Omer Tamuz. The speed of sequential asymptotic learning. *Journal of Economic Theory*, 173:383–409, 2018. doi: 10.1016/j.jet.2017.11.009.

All authors contributed equally at all stages of this project.

TABLE OF CONTENTS

INTRODUCTION

Societal outcomes are determined both organically, through the multitude of choices people make in their daily lives, and more intentionally, through political mechanisms designed to elicit input from the members of a society and coordinate behavior on the basis of that input. In both cases, outcomes are a function of a large amount of dispersed, private information. The nature and desirability of these outcomes typically depend on how, and how well, that information is aggregated.

For example, whether a macroeconomic policy will have its intended effect depends on information that is spread across many individuals; whether the intended effect of the policy is a socially good one depends on the preferences of the many individuals it will impact; and whether the choice of policy is responsive to those preferences depends on the system used for selecting a policy. How we learn from each other about individual decision-making and the institutions we use to do collective decision-making are at the heart of the complex processes that shape the society we live in.

This dissertation contains three essays. Chapters 1 and 2 provide contributions to the study of social learning, and Chapter 3 provides a contribution to the study of social choice. Chapter 1 studies a model of social learning on networks with endogenous action timing. Chapter 2, co-authored with Vadim Martynov and Omer Tamuz, studies the rate of information aggregation in the classical sequential learning model when signals are unbounded and learning occurs. Chapter 3, co-authored with Laurent Bartholdi, Maya Josyula, Omer Tamuz, and Leeat Yariv, introduces and studies a novel notion of equitability for voting rules over two alternatives.

In Chapter 1, which is based on Hann-Caruthers (2022), I introduce a fully rational model of social learning on networks with endogenous action timing. Agents receive private signals about the state of the world, which is either high or low. In the high state, agents would prefer to make a one-time irreversible investment as soon as possible, and in the low state, agents would prefer to never make the investment. At the beginning of each period $t = 0, 1, \dots,$ agents decide whether to invest or defer investment, and at the end of each period they observe the decisions made by their neighbors in that period.

I show that the structure of the network can play an important role in the aggregation of information. When the social network contains high-degree vertices, agents can be arbitrarily likely to make good choices. In contrast, when the social network is linear, there is a bound on how likely agents are to make good choices which holds regardless of how patient they are. The main contribution of this chapter is the identification of a novel mechanism through which strategic behavior can substantially impede the flow of information through a social network.

In Chapter 2, we study the asymptotic rate at which the probability of taking the correct action converges to 1 in the classical sequential learning model with unbounded signals. We show that this convergence occurs more slowly than when agents directly observe private signals. However, we also show that the associated speed of learning can be arbitrarily close to the speed of learning from signals, and that the expected time until the last incorrect action can be finite or infinite. The main contribution of this chapter is a characterization of the asymptotic law of motion of the public belief and the use of this law of motion to analyze asymptotic properties of information aggregation in the canonical observational learning model.

In Chapter 3, we introduce *equitability* as a less stringent alternative to symmetry for modeling egalitarianism in voting rules. A voting rule is equitable if all voters have the same "role" in the rule, in the sense that voters are indistinguishable in terms of the way their votes affect the final outcome. Formally, a voting rule is equitable if its group of symmetries acts transitively on the set of voters. We show that equitable voting rules can have minimal winning coalitions comprising a vanishing fraction of the population, but they cannot be smaller than the square root of the population size. The main contribution of this chapter is to introduce techniques from group theory and demonstrate their usefulness for the analysis of questions in social choice.

Chapter 1

THE ROLE OF NETWORK STRUCTURE IN THE DIFFUSION OF INVESTMENT DECISIONS

1.1 Introduction

Social networks play an important role in shaping the decisions people make. We ask friends for advice and take cues from the decisions they make, and we also understand that our friends interact with their friends in the same manner, as do our friends' friends, and so on. In recent years, there has been a large and active body of research devoted to understanding how the structure of the social network affects large-scale patterns of decision-making (for a recent survey, see, e.g., Golub and Sadler, 2017). The models used generally assume that the timing of agents' actions is determined exogenously, and so the work in this direction applies only insofar as the assumption of exogenous action timing is plausible. On the other hand, extant studies of endogenous action timing generally assume that agents all observe each other or observe some kind of summary statistic of the choices made so far, and so do not provide insights into the potential effects of network structure (see, e.g., Chamley, 2004a).

We study settings with both endogenous timing and social networks. For example, when deciding whether to see a new movie or read a new book, we often wait to see whether our friends do, and when deciding whether to make a costly investment, like putting solar panels on our home, we wait to see whether people who live near us or who we interact with regularly do so before making a decision for ourselves. Decisions like these, which concern whether to take an action which involves making some kind of irreversible investment (of time, money, etc.) but for which there is relatively little cost to deferring, are not modeled well by the assumption of exogenous timing and are often made on the basis of observations of only a subset of the population. For such decisions, the ability to defer in order to acquire information by observing others' choices naturally endogenizes action timing (see, e.g., Chamley, 2004a) and the structure of these observational interactions is intimately tied to the underlying social network.

In this paper, we demonstrate that the structure of the underlying social network can play a key role in determining the nature and quality of aggregate outcomes in a natural model of observational learning with endogenous action timing, one in which rational agents face these kinds of irreversible investment decisions. As we show, when the social network contains high-degree vertices, as when people live in a city and regularly observe choices made by many others, agents can be very likely to make good choices. In contrast, we show that when the social network is linear, as when people live along a highway in a low-population area, there is a bound on how likely agents are to make good choices, a bound which holds regardless of how patient the agents are.

This inability to efficiently aggregate information is driven by that fact that in linear networks, there are few channels for the flow of information, and rational behavior renders these channels fragile. Agents serve not only as information sources for other agents but also as information conduits, and the inefficiency is generated by their failure to internalize the impact their choices have on the communication of information through the network.

The importance of rational behavior in making the information channels fragile is underscored by the fact that this is an equilibrium phenomenon. As we show, it is possible on the line, out of equilibrium, for all agents simultaneously to be arbitrarily likely to make the right decision in the long run.

Related literature

A sizeable literature has studied learning on social networks when there is no endogeneity in action timing. When each agent acts once at an exogenously specified time, Bikhchandani et al. (1992) show that when agents see the actions of all of their predecessors, society acts suboptimally with positive probability, and similar results have been shown when the observational network is random (see, e.g., Acemoglu et al., 2011, Arieli and Mueller-Frank, 2019, Lobel and Sadler, 2015). Bala and Goyal (1998) show that when myopic agents that exhibit a certain kind of bounded rationality act repeatedly, bounded (out-)degree (together with an appropriate spread of priors) leads to learning, and Mossel et al. (2015) show that when rational agents act repeatedly, a kind of structural "egalitarianism" is sufficient for society to eventually make good decisions. Note that the line network in egalitarian, but does not allow aggregation of information in my model. Molavi et al. (2018) study long-run outcomes in models of social learning on networks across a broad class of non-Bayesian updating rules.

There has also been work on learning in settings with endogenous action timing in which the agents all observe each others' actions. The closest work to ours in this setting is Chamley (2004a), who studies a very similar base game. The author provides a characterization of symmetric equilibria in this setting and provides a nice comparison to the results in Bikhchandani et al. (1992). Several papers also explore sudden shared behavior that occurs in models of heterogeneous signal precision just after the most informed agent acts (see, e.g., Grenadier, 1999, Zhang, 1997).

1.2 Model

We consider a set of agents, N, which may be finite or countably infinite. There is a state ω, unknown to the agents, which is equally likely to be high (H) or low (L). Each agent i receives a private signal s_i about the state. We assume that the agents' signals are i.i.d. conditional on the state, and that they induce continuous and bounded beliefs; that is, we assume the induced posterior belief $p_i = \mathbb{P}(\omega = H \mid s_i)$ has a density supported on $[a, b]$ for some $0 < a < b < 1$.

There are infinitely many discrete periods $t = 0, 1, \ldots,$ and agents can make an irreversible investment in at most one period. Equivalently, agent i has to choose a time $\tau_i \in \mathbb{N} \cup \{\infty\}$ to make an investment, with $\tau_i = \infty$ corresponding to no investment. When the state is high, agents would prefer to irreversibly invest, and when the state is low, agents would prefer to never invest. Moreover, when the state is high, agents would prefer to make the investment sooner rather than later. Concretely, there is a common discount factor $\delta \in (0, 1)$, and agent i's utility is δ^{τ_i} if the state is high, $-\delta^{\tau_i}$ if the state is low, and 0 if $\tau_i = \infty$, i.e., if she never invests.

There is a social network G describing the relationships between the agents. We assume that G is bi-directional (i.e., i is j's neighbor if and only if j is i's neighbor), finite degree (i.e., each agent has only finitely many neighbors), and connected (i.e., for any agents i and j, either j is i's neighbor, or j is a neighbor of one of i's neighbors, or j is a neighbor of a neighbor of one of i's neighbors, etc.). In our main application, G will be the bi-infinite *line*, which can be thought of as the "one-dimensional" social network where there is one agent at each integer location, and each agent's neighbors are the two agents who are at a distance one from her. Another prototypical example to have in mind is the complete network, which is the social network in which every agent is neighbors with every other agent. Each agent observes the investment decisions of her neighbors at each time period.

A strategy for an agent is an investment decision (invest or do not invest) for each period t based on her private signal and the history h_t of investment decisions made by her neighbors, subject to the constraint that she can invest at most once. We assume that agents are expected utility maximizers. We will study Bayes-Nash Equilibria of this game.

Diffusion

It takes time for information to diffuse through the social network. In order for many agents to invest when the state is high and few to invest when the state is low on a general social network, it must be the case that many agents wait a long time before investing. Accordingly, we focus on what decisions agents have made after a long time. We will say that agent *i makes the right investment decision in the long run* if $\tau_i < \infty$ and $\omega = H$, or if $\tau_i = \infty$ and $\omega = L$. That is, agent i makes the right investment decision in the long run if she eventually invests in the high state and never invests in the low state.

On a finite social network, there is a trivial bound on how likely any agent is to make the right investment decision in the long run, simply because the agents' decisions are made on the basis of finitely many private signals. However, for social networks that do not have some kind of structural impediment to the flow of information, it should be possible, when the population is large and agents are patient, for agents to be arbitrarily likely to make the right investment decision in the long run. We would like to understand when such structural impediments do exist. Accordingly, we will say that a family of finite networks (G_n) *obstructs diffusion* if there is a $\bar{p} < 1$ such that every agent's probability of making the right investment decision in the long run in every equilibrium on every G_n for every value of the discount factor δ is at most $\bar{p}$.

On an infinite social network, it is in principle possible for an agent to be arbitrarily likely to make the right investment decision in the long run, and so it is reasonable to apply this criterion directly to a single network in this case: we will say that an infinite social network G obstructs diffusion if there is a $\bar{p} < 1$ such that every agent's probability of making the right investment decision in the long run in every equilibrium on G for every value of the discount factor δ is at most $\bar{p}$.

1.3 Results

Say that a family of finite networks is of unbounded degree (G_n) if for every degree d, there is at least one agent in at least one network G_n who has at least d neighbors. Our first result is that families of unbounded degree do not obstruct diffusion.

Theorem 1. *If (G_n) is a family of finite networks of unbounded degree, then (G_n) does not obstruct diffusion.*

Chamley (2004a) proves this for the family of complete networks (K_n). Following the conventions used in that paper, denote by $\bar{\mu}$ the highest possible belief induced by a private signal, and denote by μ^{**} the belief at which an agent would be indifferent between investing in period 0 and deferring, assuming that the state will be perfectly revealed to her between periods 0 and 1.

Chamley shows that for values of δ such that $\mu^{**} < \bar{\mu}$, the first period symmetric equilibrium strategy is unique for large enough n and information is asymptotically perfect at the end of period 0. It follows that when μ^{**} is less than but very close to $\bar{\mu}$, each agent is very likely to defer in period 0. Nevertheless, observing the number of agents that do invest yields a very informative signal about the state. Then, with high probability, agents invest in period 1 if and only if the state is high.

We extend this result to general families of finite networks of unbounded degree (G_n). In the same spirit, consider values of δ so that μ^{**} is less than but very close to $\bar{\mu}$. In equilibrium, agents will always invest in period 0 if their belief is above μ^{**} and will never invest in period 0 if their belief is below $\frac{1}{2}$, so each agent's period 0 decision is informative.

Now, let (G_{n_j}) be a sequence of networks and let (i_j) be a sequence of agents of growing degree, so that the number of neighbors agent i_j has in G_{n_j} approaches infinity as j approaches infinity. Taken together, the period 0 decisions of agent i_j's neighbors become arbitrarily informative as j grows, so i_j's information is asymptotically perfect at the end of period 0. Moreover, i_j's probability of investing in period 0 becomes vanishingly small as j tends to infinity, and as we vary δ to make μ^{**} approach $\bar{\mu}$.

Thus, for any $\epsilon > 0$, it is possible to find an agent i_j, a $\delta < 1$ and an equilibrium on G_{n_j} so that with probability at least $1 - \epsilon$, i_j defers in period 0, then invests in period 1 if the state is high and never invests if the state is low. This implies, in particular, that there is no universal bound on the probability of agents making the right investment decision (eventually invest in the high state, never invest in the low state) in the long run in equilibrium.

Our main result is that, in contrast to families of unbounded degree, such a universal bound does exist for equilibria on the line. Recall that the line is the social network where the agents are identified with the set of integers $\mathbb{Z}$, and two agents i and j are neighbors if and only if they are one apart, $|i - j| = 1$.

Theorem 2. *The line obstructs diffusion.*

At the heart of Theorem 2 is the fact that agents serve not only as sources of information for other agents, but also as media for the flow of information. They do not take into account the impact of their decisions on the information channels they are a part of, and because there are very few information channels, they create blockages which cannot be circumvented. In many classical learning models, the problem is that only a small amount of information enters the system through the agents' actions. In contrast, in this model, a hypothetical outside observer, who could see the decisions made by all of the agents, would learn the state by observing just the actions taken in period 0. As we explore in the next section, the problem is how well this information is communicated across the network.

1.4 Equilibria on the line

To give a proper account of forces underpinning Theorem 2, it is important to have a sense of the structure of equilibria on the line. Concisely, from the perspective of an outside observer who sees only the decisions the agents make, equilibrium outcomes look very much like outcomes in standard infection or diffusion models: some agents invest immediately (are initially infected), then the decision to invest (the infection) "spreads" some distance in either direction from the immediate investors until reaching agents who do not invest (until transmission fails).

Note that these dynamics are not assumed; they necessarily arise endogenously in equilibrium. To see why outcomes take this form, observe that agents face an optimal stopping problem in which they prefer to invest sooner rather than later in the high state and to never invest in the low state, and so will be more inclined to invest in any period the higher their belief is, all else equal. Consequently, agents use threshold rules in equilibrium: at each history where she is considering whether to invest, agent i has a threshold belief such that she invests at that history if and only if her belief is above that threshold. So in each period, agents who have not yet invested do so if they are optimistic enough about the prospects of investing and defer in order to gather more information if they are not so optimistic.

Intuitively, one would expect that an agent who has decided to postpone would not, after a period of seeing no investment activity from her neighbors who also postponed, suddenly determine that it is worth investing–her neighbors' continued lack of investment should only make her more pessimistic about the prospects of investing. On the line, this intuition is correct:[1] in equilibrium, agents do not invest in any period that immediately follows a period of inactivity from their neighbors. More concretely, if none of an agent's neighbors invests in period t, then she does not invest in period $t + 1$.[2] This entails the infection-style outcomes described above: some agents who receive encouraging private signals, invest in period 0, the neighbors of the period 0 investors who did not invest in period 0 but who received sufficiently encouraging private signals, then adopt in period 1, and so on.

Clogging the information channels

With this picture in mind, consider the choice facing agent i in period t, assuming she did not invest until then. Suppose neither of her neighbors invested in periods $0, \ldots, t - 2$ and exactly one of her neighbors, say agent $i - 1$, invested in period $t - 1$. From agent $i - 1$'s investment choice, agent i infers that agent $i - t$ invested in period 0, agent $i - (t - 1)$ invested in period 1, and so on.

If agent i decides to invest, she effectively passes on this information to agent $i + 1$, along with some information about her own private signal. When this happens, the information $i + 1$ receives from agent i bears some resemblance to the information an agent receives in the classical sequential learning model, with agents $i - t$ through i playing the role of agent $i + 1$'s predecessors.

However, if she chooses not to invest, she passes very little of this information to agent $i + 1$, and the similarity to the classical sequential learning model becomes weaker. In this case, agent $i + 1$ learns relatively little from observing that agent i chooses not to invest in period t.

Now, it is possible that in such a scenario, agent i chooses not to invest in period t because she is genuinely pessimistic about the prospects of investing. In an analogous observational learning model where agents act once at an exogenously determined time, this would in fact be the only reason an agent would not invest.

However, agent i's decision about whether to invest at this point is driven not only by her belief, but also by her expectations about the information that will be available

[1] In fact, this intuition is correct whenever the social network is a tree. We conjecture that it is also correct on a general social network, but the methods we use do not immediately apply when the network is not a tree.

[2] This does not mean that an agent necessarily invests in period $t + 1$ if some of her neighbors invest in period t.

to her in the future. In particular, it is possible that agent i is reasonably confident that the state is high, but that she nevertheless defers because she expects that by observing agent $i+1$'s future decisions, she will get strong enough information soon enough and will then be able to make a more informed decision.

This means that the stronger an information source agent $i + 1$ is for agent i, the stronger the incentive is for agent i to defer investment, and the weaker her proclivity to take actions that are informative for agent $i + 1$. Put another way, the more conducive agent $i + 1$'s behavior is to the flow of information, the stronger the inducement is to agent i to obstruct this flow. Hence, there is a limit on how well the agents perform in the long run.

Physical impediments to the flow of information

Even without the equilibrium pressure to clog the information channels, one might suspect that Theorem 2 is a consequence of the fact that communication channels on the line are already very fragile–as soon as an agent invests or decides she will never invest, she cuts off the flow of information between the agents on either side of her. However, as the following result illustrates, Theorem 2 is not the consequence of some kind of "physical" obstruction. It is the strategic aspects of the model that prevent agents on the line from getting and making use of a large amount of information.

Proposition 1. *For any $\epsilon > 0$, there is a symmetric strategy profile on the line such that every agent's probability of making the right investment decision in the long run is at least $1 - \epsilon$.*

In fact, something similar holds for any large or infinite network. The key observation underlying this result is that, because agents can choose not only whether to invest but also a time at which to invest, there is no bound on how much information can be communicated between neighbors. For example, if agent i wanted to communicate a very good approximation of her belief to agent $i + 1$, the agents could in principle agree on a protocol for when agent i should invest depending on her signal.

If one imagines such a protocol for aggregating the signals of agents 1 through k, all agents $j > k$ could make the same investment decision on the basis of this same information, and if k is large, they would all make the right decision in the long run with very high probability. The symmetric strategy profile in Proposition 1 essentially combines this protocol with a randomization for each agent about what role they play they in the protocol.

One implication of Proposition 1 is that, for patient agents, it is possible to approximate the first-best welfare outcome. Note that under any strategy profile, each agent's expected payoff is at most $\frac{1}{2}$, since each agent can get a payoff of at most 0 in the low state and 1 in the high state. Under the strategy profile in Proposition 1, the expected payoff to every agent nearly achieves this upper bound.

Corollary 1. *For any $\epsilon > 0$, there is a $\delta < 1$ and a symmetric strategy profile on the line under which every agent's expected payoff is at least $\frac{1}{2} - \epsilon$.*

1.5 Conclusion

In this paper, we have shown that the network structure can have a profound influence on aggregate outcomes when timing is endogenous. In particular, in some networks the agents' strategic behavior leads them to obstruct the flow of information, which results in limited information aggregation, while in others there is no such obstruction.

The model studied in this paper is rich, and we expect that we have only scratched the surface. Studying the influence of particular features of the network on strategic behavior has the potential to yield insights about the relationship between rational behavior and diffusion processes that are not within the scope of other models of social learning on networks.

For example, we conjecture that the three-regular tree, the infinite tree network in which every agent has three neighbors, does not obstruct diffusion. Such a network is far from having unbounded degree, but still has the potential to give rise to good outcomes in the long run because it has many information channels.

In a different direction, it would be especially interesting to find sufficient structural conditions for when a family of finite networks obstructs diffusion. For example, our proof for obstruction on the infinite line applies to finite lines, but not to finite cycles, for which additional ideas will be needed.

1.6 Setting up the optimal stopping rule problem

For analysis, it will be useful to take the space of private signals for agent i to be $[0, 1]$ and to take the signal s_i to be the induced posterior, i.e., $s_i = \mathbb{P}(\omega = H \mid s_i)$. This is without loss of generality, because this belief is a sufficient statistic for ω given the signal. It will be likewise useful to think of τ_i, the strategy of agent i, as a stopping rule with respect to her private signal s_i and her neighbors' investment times. That is, we will think of agent i's strategy as a function τ_i taking values in $\mathbb{N} \cup \{\infty\}$ such that $\mathbb{1}(\tau_i = t)$ can depend only on whether the investment times of i's neighbors are less than t and the particular values of those investment times for the neighbors whose investment times are less than t.

Observe that, given the strategies of the other agents, agent i's problem is to find an optimal stopping rule τ_i. If agent i adopts at time t, it means that she observed how her neighbors behaved in periods 0 through $t-1$ in response to her not investing in those periods; in particular, it means that any actions her neighbors take after she invests do not factor into any of her decisions. Hence, agent i's problem is to find an optimal stopping rule with respect to the stream of signals produced by her neighbors' actions contingent on her not investing.

Given a strategy profile $\tau = (\tau_i)$, denote by h_t^i the history of actions taken in periods 0 through $t-1$ by i's neighbors. Denote by $\tilde{h}_t^i$ the history of actions taken by i's neighbors under the strategy profile (σ_i, τ_{-i}) where σ_i is the strategy under which i never invests. Since i effectively exits the game once she invests, given her private signal s_i, i's problem is to find an optimal stopping rule with respect to the signal stream $(\tilde{h}_1^i, \tilde{h}_2^i, \ldots)$.

1.7 Optimal stopping rules are always threshold strategies

Given any history h_t^i and any signal s_i, either agent i has already invested by time t, or agent i is determining whether she should invest in period t or defer. In the latter case, she is deciding whether stopping immediately (investing in period t) is an optimal stopping rule, given the information she has so far. For any stopping time τ, define

$$V(\tau, t;\, \omega, h_t^i) := \mathbb{E}(\delta^t - \delta^\tau \,|\, \omega,\, h_t^i).$$

Then we have:

Lemma 1. *It is optimal for agent i to invest at the history h_t^i given her signal s_i if and only if*

$$\mathbb{P}(\omega = H \,|\, h_t^i,\, s_i) \cdot V(\tau, t;\, H, h_t^i) \geq \mathbb{P}(\omega = L \,|\, h_t^i,\, s_i) \cdot V(\tau, t;\, L, h_t^i)$$

for all stopping times τ with respect to $(\tilde{h}_{t+1}^i, \tilde{h}_{t+2}^i, \ldots)$ given the history h_t^i.

We will say that agent i is using a *threshold strategy* if for any history h_t^i, there is some s such that given h_t^i, $\tau_i \leq t$ if and only if $s_i \geq s$.

Proposition 2. *Best responses are always almost surely threshold strategies on path.*

Proof. If not, there would be some history where a threshold strategy is not used. But by the lemma, this would mean that the behavior is suboptimal with nonzero probability, contradiction. □

1.8 Families of unbounded degree

Proof of Theorem 1. Let $\bar{\mu}$ be the highest possible belief induced by a private signal, and for any discount factor δ, let $\mu^{**}(\delta)$ be defined by

$$1 \cdot \mu^{**}(\delta) + -1 \cdot (1 - \mu^{**}(\delta)) = \delta(1 \cdot \mu^{**}(\delta) + 0 \cdot (1 - \mu^{**}(\delta))).$$

Note that $\mu^{**}(\delta) \to \frac{1}{2}$ as $\delta \to 0$ and $\mu^{**}(\delta) \to 1$ as $\delta \to 1$.

Note that if $s_i \geq \mu^{**}(\delta)$, investing immediately is strictly dominant for agent i, and if $s_i < \frac{1}{2}$, deferring in period 0 is strictly dominant.

So when the discount factor is δ, in any equilibrium on any network, every agent's threshold for investing in period 0 must be between $\frac{1}{2}$ and $\mu^{**}(\delta)$. There is a lower bound $e > \frac{1}{2}$ on the probability of correctly identifying the state on the basis of $\mathbb{1}(s_i \geq \underline{s})$ that holds for all $\frac{1}{2} \leq \underline{s} \leq \mu^{**}(\delta)$. It follows that e is also a lower bound on the probability of correctly identifying the state on the basis of any single agent's period 0 decision. Hence, because the agents' period 0 decisions are independent conditional on the state, the probability of correctly identifying the state on the basis of the period 0 decisions of any k agents approaches 1 as $k \to \infty$.

Denote by $\bar{\delta}$ the solution to $\mu^{**}(\delta) = \bar{\mu}$. Consider the strategy of deferring in period 0, investing in period 1 if her belief is above $\frac{1}{2}$, and never investing otherwise. It follows that the expected payoff from using this strategy approaches $\frac{1}{2}\bar{\delta}$ as the number of neighbors an agent has grows and δ approaches $\bar{\delta}$.

It follows that for any $\epsilon > 0$ and any agent i with a large enough number of neighbors in some G_n, there is a δ so that i's expected payoff in any equilibrium is at least $\frac{1}{2}\bar{\delta} - \epsilon$. This is only possible if the probability that the agent invests in period 1 in the high state approaches 1 and the probability that the agent eventually invests in the low state approaches 0. □

1.9 No spontaneous investment on trees

We will say that a strategy σ_i for agent i satisfies *no spontaneous investment* if for every $t \geq 1$ and every history h_t^i in which none of her neighbors invests in period $t-1$, agent i does not invest in period t, and we will say that a strategy profile satisfies no spontaneous investment if every agent's strategy satisfies no spontaneous investment.

Proposition 3. *If G is a tree, then every equilibrium satisfies no spontaneous investment.*

A simple but useful observation is that agents never invest when they believe that the state is more likely to be low than high.

Lemma 2. *In any equilibrium, if $\mathbb{P}(\omega = H \mid h_t^i, s_i) < \mathbb{P}(\omega = L \mid h_t^i, s_i)$, then agent i does not invest in period t.*

Another useful fact is that if G is a tree, then seeing a neighbor invest is always evidence that the state is high.

Lemma 3. *Suppose G is a tree and we have an equilibrium. For any i and j, let τ_j^i be the investment time for agent j under the strategy profile where all agents except i keep the same strategy and agent i uses the never invest strategy. If i and j are neighbors, then*

$$\mathbb{P}(\tau_j^i = t \mid \omega = H) \geq \mathbb{P}(\tau_j^i = t \mid \omega = L).$$

An important corollary is that when G is a tree, seeing more neighbors invest is always stronger evidence that the state is high.

Lemma 4. *Suppose G is a tree and we have an equilibrium. Let s_i be a private signal and h_{t+1}^i and $h_{t+1}^{i}{}'$ histories such that i does not invest before $t+1$ under either history, such that $h_t^i = h_t^{i\prime}$, and such that the set of i's neighbors who invest in period t in h_{t+1}^i is a subset of the set of neighbors who invest in period t in $h_{t+1}^{i}{}'$. Then*

$$\mathbb{P}(\omega = H \mid h_{t+1}^{i}{}', s_i) \geq \mathbb{P}(\omega = H \mid h_{t+1}^i, s_i).$$

Proof. Let S be the set of i's neighbors who invest at t in ${h^i_{t+1}}'$ but not h^i_{t+1}. Then

$$\begin{aligned}
&\frac{\mathbb{P}(\omega = H \mid {h^i_{t+1}}', s_i)/\mathbb{P}(\omega = L \mid {h^i_{t+1}}', s_i)}{\mathbb{P}(\omega = H \mid h^i_{t+1}, s_i)/\mathbb{P}(\omega = L \mid h^i_{t+1}, s_i)} \\
&= \prod_{j \in S} \frac{\mathbb{P}(\tau^i_j = t \mid \omega = H)/\mathbb{P}(\tau^i_j = t \mid \omega = L)}{\mathbb{P}(\tau^i_j > t \mid \omega = H)/\mathbb{P}(\tau^i_j > t \mid \omega = L)} \\
&= \prod_{j \in S} \frac{\mathbb{P}(\tau^i_j = t \mid \omega = H)}{\mathbb{P}(\tau^i_j = t \mid \omega = L)} \cdot \frac{\mathbb{P}(\tau^i_j > t \mid \omega = L)}{\mathbb{P}(\tau^i_j > t \mid \omega = H)} \\
&\geq 1.
\end{aligned}$$

□

Proof of Proposition 3. Suppose not. Then there is some history h^i_{t+1} where none of agent i's neighbors invests in period t but i invests in period $t+1$.

First, note that if given h^i_t, agent i adopts in period $t+1$ regardless of the actions her neighbors take in period t, then by Lemma 2 she must find the high state at least as likely as the low state in all of these cases, and so she must find the high state at least as likely in the low state at the history h^i_t. That is, $\mathbb{P}(\omega = H \mid h^i_t, s_i) \geq \mathbb{P}(\omega = L \mid h^i_t, s_i)$. But then her expected utility given h^i_t is

$$\delta^{t+1}(\mathbb{P}(\omega = H \mid h^i_t, s_i) - \mathbb{P}(\omega = L \mid h^i_t, s_i)).$$

On the other hand, if she invests at h^i_t, her expected utility given h^i_t is

$$\delta^{t}(\mathbb{P}(\omega = H \mid h^i_t, s_i) - \mathbb{P}(\omega = L \mid h^i_t, s_i)),$$

which is greater.

Thus, it must be that there is some subset S of her neighbors such that, given h^i_t, if all (and only) agents in S invest in period t, then i does not invest in period $t+1$. Denote by ${h^i_{t+1}}'$ this history. By Lemma 1, there is some τ such that

$$\mathbb{P}(\omega = H \mid {h^i_{t+1}}', s_i) \cdot V(\tau, t;\, H, {h^i_{t+1}}') < \mathbb{P}(\omega = L \mid {h^i_{t+1}}', s_i) \cdot V(\tau, t;\, L, {h^i_{t+1}}').$$

Note that τ depends only on the investment decisions of the neighbors of i who are not in S and who do not invest before t under h^i_t. Thus, τ is also a well-defined stopping time starting from h^i_{t+1}. Moreover, because G is a tree, the investment decisions of these agents are independent of the history of investment decisions of i's other neighbors given that i has not invested, so $V(\tau, t;\, H, {h^i_{t+1}}') = V(\tau, t;\, H, h^i_{t+1})$. By Lemma 4, $\mathbb{P}(\omega = H \mid h^i_t,\, s_i) \le \mathbb{P}(\omega = H \mid {h^i_{t+1}}',\, s_i)$. Hence,

$$\mathbb{P}(\omega = H \mid h^i_{t+1},\, s_i) \cdot V(\tau, t;\, H, h^i_{t+1}) < \mathbb{P}(\omega = L \mid h^i_{t+1},\, s_i) \cdot V(\tau, t;\, L, h^i_{t+1}),$$

so by Lemma 1, investing at h^i_{t+1} is strictly suboptimal, contradiction. □

1.10 Making the right investment decision at infinity as a spectator

Define agent i's *probability of making the right investment decision at infinity as a spectator* p_i^∞ to be the probability that the agent would make the right investment decision given s_i and τ^i_j for $j \in N_G(i)$. It follows easily from the definition that i's probability of making the right investment decision in the long run is bounded above by her probability of making the right investment decision at infinity as a spectator, since at infinity she can mimic her strategy (that is, invest at infinity if $\tau_i < \infty$ and do not invest at infinity if $\tau_i = \infty$).

Correspondingly, define agent i's *spectator belief at infinity*

$$\pi_i^\infty = \mathbb{P}(\omega = H \mid s_i, (\tau^i_j)_{j \in N_G(i)}).$$

In order for agent i to be very likely to make the right investment decision, it must be that information she has available is very strong. In particular, it must be that π_i^∞ is almost always very close to 1 in the high state. This yields a sufficient condition for a network to obstruct diffusion:

Proposition 4. *If there exist $p > 0$ and $x < 1$ such that*

$$\mathbb{P}(\pi_i^\infty \le x \mid \omega = H) \ge p$$

for every agent i under every equilibrium on G, then G obstructs diffusion.

Proof. Without loss of generality, we can assume $x > \frac{1}{2}$. First, note that

$$\frac{\mathbb{P}\left(\frac{1}{2} < \pi_i^\infty \le x \mid \omega = L\right)}{\mathbb{P}\left(\frac{1}{2} < \pi_i^\infty \le x \mid \omega = H\right)} = \frac{\mathbb{P}\left(\omega = L \mid \frac{1}{2} < \pi_i^\infty \le x\right)}{\mathbb{P}\left(\omega = H \mid \frac{1}{2} < \pi_i^\infty \le x\right)} > \frac{1-x}{x}.$$

Hence

$$\begin{aligned}1 - p_i^\infty &= \mathbb{P}(\omega = L) \cdot \mathbb{P}\left(\pi_i^\infty > \frac{1}{2} \,|\, \omega = L\right) + \frac{1}{2}\mathbb{P}\left(\pi_i^\infty = \frac{1}{2}\right) + \mathbb{P}(\omega = H) \cdot \mathbb{P}\left(\pi_i^\infty < \frac{1}{2} \,|\, \omega = H\right)\\ &\geq \frac{1}{2} \cdot \mathbb{P}\left(\frac{1}{2} < \pi_i^\infty \leq x \,|\, \omega = L\right) + \frac{1}{2}\mathbb{P}\left(\pi_i^\infty = \frac{1}{2}\right) + \frac{1}{2} \cdot \mathbb{P}\left(\pi_i^\infty < \frac{1}{2} \,|\, \omega = H\right)\\ &\geq \frac{1}{2} \cdot \frac{1-x}{x} \cdot \mathbb{P}\left(\frac{1}{2} < \pi_i^\infty \leq x \,|\, \omega = H\right) + \frac{1}{2}\mathbb{P}\left(\pi_i^\infty = \frac{1}{2}\right) + \frac{1}{2} \cdot \mathbb{P}\left(\pi_i^\infty < \frac{1}{2} \,|\, \omega = H\right)\\ &\geq \frac{1}{2} \cdot \frac{1-x}{x} \cdot \mathbb{P}(\pi_i^\infty \leq x \,|\, \omega = H)\\ &\geq \frac{1}{2} \cdot \frac{1-x}{x} \cdot p,\end{aligned}$$

so the probability that agent i makes the right investment decision in the long run is at most $1 - \frac{1}{2} \cdot \frac{1-x}{x} \cdot p$. □

1.11 Dynamics on the line

As a reminder, the agents' private signals induce bounded beliefs. In particular, there is some $\alpha > 1$ such that

$$\frac{1}{\alpha} \leq \frac{s_i}{1 - s_i} \leq \alpha$$

almost surely.

We will need the following fact.

Lemma 5. *Let $A, B \subseteq [0,1]$, with $\mathbb{P}(s_i \in A \cap B \,|\, \omega = H) > 0$. If $\mathbb{P}(s_i \in A \,|\, s_i \in B,\ \omega = H) \geq \frac{1}{2}$, then*

$$\mathbb{P}(s_i \in A \,|\, s_i \in B,\ \omega = L) \geq \mathbb{P}(s_i \in A \,|\, s_i \in B,\ \omega = H)^{2\alpha^2}.$$

In particular, we will need the following corollary.

Lemma 6. *Let $S \subseteq N$ be a finite set of agents, and let $A_i, B_i \subseteq [0,1]$ for each $i \in S$. Let A be the event that $s_i \in A_i$ for all $i \in S$, and let B be the event that $s_i \in B_i$ for all $i \in S$. Suppose that $\mathbb{P}(A \cap B \,|\, \omega = H) > 0$. If $\mathbb{P}(A \,|\, B, \omega = H) \geq \frac{1}{2}$, then*

$$\mathbb{P}(A \,|\, B, \omega = L) \geq \left(\frac{1}{2}\right)^{2\alpha^2}.$$

Now, let $C_{i,t}$ be the event that agents $i+1, \ldots, i+t$ do not invest in period 0 and agent $i+t+1$ does.

Lemma 7. *If* $\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = H) \geq \frac{1}{2}$*, then* $\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = L) \geq \left(\frac{1}{2}\right)^{2\alpha^2}$.

We have the following useful fact:

Proposition 5. *There is a* $\kappa > 0$ *such that for any infinite G, any* δ*, and any equilibrium with respect to* δ *that satisfies no spontaneous investment,*

$$\mathbb{P}(\tau_j > 0 \text{ for all } j \in B_G(i,r)) \leq \delta^{\kappa r}$$

for all sufficiently large r for every $i \in N$*.*

It follows that $\sum_{t\geq 0} \mathbb{P}(C_{i,t}) = 1$.

We are now ready to show that agent $i+1$'s decisions cannot be arbitrarily likely to be arbitrarily informative for agent i.

Proposition 6. *There is a* $p > 0$ *and an* $x < 1$ *such that for any equilibrium on the line,*

$$\mathbb{P}(\mathbb{P}(\omega = H \mid \tau^i_{i+1}) \leq x \mid \omega = H) \geq p.$$

Proof. Let $x = \frac{\alpha 2^{2\alpha^2}}{1+\alpha 2^{2\alpha^2}}$ and $p = \frac{1}{2}$. Let $Z = \{t : \mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = H) \geq \frac{1}{2}\}$.

For any $t \in Z$,

$$\begin{aligned}
\frac{\mathbb{P}(\omega = H \mid \tau^i_{i+1} = t)}{\mathbb{P}(\omega = L \mid \tau^i_{i+1} = t)} &= \frac{\mathbb{P}(\tau^i_{i+1} = t \mid \omega = H)}{\mathbb{P}(\tau^i_{i+1} = t \mid \omega = L)} \\
&= \frac{\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = H)\mathbb{P}(C_{i,t} \mid \omega = H)}{\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = L)\mathbb{P}(C_{i,t} \mid \omega = L)} \\
&\leq 2^{2\alpha^2} \frac{\mathbb{P}(C_{i,t} \mid \omega = H)}{\mathbb{P}(C_{i,t} \mid \omega = L)}.
\end{aligned}$$

Now, since agents use threshold strategies, they are more likely to invest immediately in the high state, so

$$\frac{\mathbb{P}(C_{i,t} \mid \omega = H)}{\mathbb{P}(C_{i,t} \mid \omega = L)} \leq \frac{\mathbb{P}(\tau_{i+t+1} = 0 \mid \omega = H)}{\mathbb{P}(\tau_{i+t+1} = 0 \mid \omega = L)} \leq \alpha.$$

Thus,

$$\frac{\mathbb{P}(\omega = H \mid \tau^i_{i+1} = t)}{\mathbb{P}(\omega = L \mid \tau^i_{i+1} = t)} \leq \alpha 2^{2\alpha^2},$$

so

$$\mathbb{P}(\omega = H \mid \tau_{i+1}^i = t) \le \frac{\alpha 2^{2\alpha^2}}{1 + \alpha 2^{2\alpha^2}} = x.$$

Further, it follows from another lemma way earlier that

$$\frac{\mathbb{P}(\omega = H \mid \tau_{i+1}^i = \infty)}{\mathbb{P}(\omega = L \mid \tau_{i+1}^i = \infty)} = \frac{\mathbb{P}(\tau_{i+1}^i = \infty \mid \omega = H)}{\mathbb{P}(\tau_{i+1}^i = \infty \mid \omega = L)} \le 1 < x.$$

Hence, $\mathbb{P}(\omega = H \mid \tau_{i+1}^i = t) > x$ implies $t \notin Z$, so

$$\begin{aligned}
\mathbb{P}(\mathbb{P}(\omega = H \mid \tau_{i+1}^i) > x \mid \omega = H) &= \sum_t \mathbb{P}(\mathbb{P}(\omega = H \mid \tau_{i+1}^i) > x \mid C_{i,t}, \omega = H)\mathbb{P}(C_{i,t} \mid \omega = H) \\
&\le \sum_{t \notin Z} \mathbb{P}(\tau_{i+1}^i = t \mid C_{i,t}, \omega = H)\mathbb{P}(C_{i,t} \mid \omega = H) \\
&< \sum_{t \notin Z} \frac{1}{2}\mathbb{P}(C_{i,t} \mid \omega = H) \\
&< \frac{1}{2}.
\end{aligned}$$

Thus, $\mathbb{P}(\mathbb{P}(\omega = H \mid \tau_{i+1}^i) \le x \mid \omega = H) \ge p$. □

By completely symmetric arguments, it also holds that agent $i-1$'s decisions cannot be arbitrarily likely to be arbitrarily informative for agent i.

We are now ready to prove our main theorem.

Theorem 3. *The line obstructs diffusion.*

Proof. Let p and x be as in the previous proposition. Then

$$\begin{aligned}
\frac{\mathbb{P}(\omega = H \mid \tau_{i-1}^i, \tau_{i+1}^i, s_i)}{\mathbb{P}(\omega = L \mid \tau_{i-1}^i, \tau_{i+1}^i, s_i)} &= \frac{\mathbb{P}(\tau_{i-1}^i, \tau_{i+1}^i, s_i \mid \omega = H)}{\mathbb{P}(\tau_{i-1}^i, \tau_{i+1}^i, s_i \mid \omega = L)} \\
&= \frac{\mathbb{P}(\tau_{i-1}^i \mid \omega = H)}{\mathbb{P}(\tau_{i-1}^i \mid \omega = L)} \frac{\mathbb{P}(\tau_{i+1}^i \mid \omega = H)}{\mathbb{P}(\tau_{i+1}^i \mid \omega = L)} \frac{\mathbb{P}(s_i \mid \omega = H)}{\mathbb{P}(s_i \mid \omega = L)},
\end{aligned}$$

so $\mathbb{P}(\tau_{i-1}^i \mid \omega = H) \le x$ and $\mathbb{P}(\tau_{i+1}^i \mid \omega = H) \le x$ implies $\frac{\pi_i^\infty}{1-\pi_i^\infty} \le \alpha \left(\frac{x}{1-x}\right)^2$.

Hence,

$$\mathbb{P}(\pi_i^\infty \le \frac{\alpha x^2}{\alpha x^2 + (1-x)^2} \mid \omega = H) \ge \mathbb{P}(\mathbb{P}(\tau_{i-1}^i \mid \omega = H) \le x, \mathbb{P}(\tau_{i+1}^i \mid \omega = H) \le x \mid \omega = H) \ge p^2.$$

The result then follows from that other proposition. □

1.12 Outside observer learns from period 0 decisions

Proposition 7. *Let p^O be the belief of an outside observer after seeing the agents' period* 0 *decisions. For any equilibrium for any value of δ on the line,* $\mathbb{P}(p^O = 1 \mid \omega = H) = 1$ *and* $\mathbb{P}(p^O = 0 \mid \omega = L) = 1$.

Proof. By Proposition 5, there is an r such that in any ball of radius r, the probability that at least one agent invests in period 0 is at least $\frac{1}{2}$. Since there is a lower bound on the ratio of how likely this is in the high state versus the low state, there is a corresponding lower bound on the divergence of the distribution of the period 0 decisions in any ball of radius r in the high state from the distribution in the low state, it follows that the state is identified perfectly from the outcomes in infinitely many mutually disjoint balls of radius r. □

1.13 No physical impediments on the line

Proof of Proposition 1. Let $\underline{s} > \frac{1}{2}$ be such that $\mathbb{P}(s_i > \underline{s} \mid \omega = H) > 0$.

Define x_i to be 0 if $s_i < \frac{1}{2}$, 1 if $\frac{1}{2} \leq s_i \leq \underline{s}$, and 2 if $s_i > \underline{s}$.

Define $q(\omega) = \frac{\mathbb{P}(x_i=1 \mid \omega)}{\mathbb{P}(x_i=0 \mid \omega)+\mathbb{P}(x_i=1 \mid \omega)}$.

Fix $k \in \mathbb{N}$, and define a corresponding strategy as follows. If $x_i = 2$, invest in period 0. If $x_i \neq 2$ and agent $i-1$ invests in period $m(k+1)+j$, where $m, j \leq k$, invest in period $(m+1)(k+1)+j$ if $x_i = 0$ and invest in period $(m+1)(k+1)+j+1$ if $x_i = 1$. If agent $i-1$ invests in period $(k+1)^2 + j$, invest in period $(k+2)(k+1)$ if

$$\left(\frac{q(H)}{q(L)}\right)^j \cdot \left(\frac{1-q(H)}{1-q(L)}\right)^{k-j} \geq \frac{1}{2}.$$

If agent $i-1$ invests in any period $t \geq (k+2)(k+1)$, invest in period $t+1$. In all other cases, never invest.

Take $\mathbb{P}(s_i > \underline{s} \mid \omega = H) \to 0$ and $k \to \infty$ in such a way that $\mathbb{P}(s_i \leq \underline{s} \mid \omega = H)^k \to 1$. Then probability that agent i makes her decision on the basis of k independent draws of an informative signal approaches 1, and the probability that she makes the right decision on the basis of these signals approaches 1. Hence, the probability that she makes the right decision in the long run approaches 1. □

1.14 Proofs of lemmas

Proof of Lemma 1. The expected payoff from any admissible stopping time τ is

$$\mathbb{E}(\mathbb{1}(\omega = H)\delta^\tau - \mathbb{1}(\omega = L)\delta^\tau \mid h_t^i, s_i).$$

Hence, stopping immediately is optimal if and only if

$$\mathbb{E}(\mathbb{1}(\omega = H)[\delta^t - \delta^\tau] \mid h_t^i,\ s_i) \geq \mathbb{E}(\mathbb{1}(\omega = L)[\delta^t - \delta^\tau] \mid h_t^i,\ s_i)$$

for all admissible τ. Since τ is independent of s_i given ω,

$$\mathbb{E}(\mathbb{1}(\omega)[\delta^t - \delta^\tau] \mid h_t^i,\ s_i) = \mathbb{P}(\omega \mid h_t^i,\ s_i) \cdot V(\tau, t;\ \omega, h_t^i).$$

Since $V(\tau, t;\ \omega, h_t^i) \geq 0$, the claim follows. □

Proof of Lemma 2. Let $\tau \equiv \infty$. Then $V(\tau, t;\ \omega, h_t^i) = \delta^t$, and since

$$\mathbb{P}(\omega = H \mid h_t^i,\ s_i)\delta^t < \mathbb{P}(\omega = L \mid h_t^i,\ s_i)\delta^t,$$

it follows from Lemma 1 that it is not optimal for i to invest at h_t^i. □

Proof of Lemma 3. If $t = 0$, this follows from Lemma 2, since every agent is at least as likely to invest in period 0 in the high state as in the low state.

Assume the claim holds for all $t' < t$, and suppose that for some signal s_j, agent j invests at a history h_t in which agent i does not invest before t. Then by Lemma 2,

$$\mathbb{P}(\omega = H \mid h_t^j, s_j) \geq \mathbb{P}(\omega = L \mid h_t^j, s_j)$$

under the equilibrium profile. Since j does not invest before t, this implies that

$$\mathbb{P}(\omega = H \mid (\tau_k^j|_t), s_j) \geq \mathbb{P}(\omega = L \mid (\tau_k^j|_t), s_j),$$

where k ranges across j's neighbors, and hence

$$\mathbb{P}((\tau_k^j|_t), s_j \mid \omega = H) \geq \mathbb{P}((\tau_k^j|_t), s_j \mid \omega = L).$$

It follows from the assumption that

$$\mathbb{P}(\tau_i^j \geq t \mid \omega = H) \leq \mathbb{P}(\tau_i^j \geq t \mid \omega = L),$$

and thus, since the $\tau_k^j|_t$ are conditionally independent because G is a tree,

$$\mathbb{P}((\tau_k^j|_t)_{k \neq i}, s_j \mid \omega = H) \geq \mathbb{P}((\tau_k^j|_t)_{k \neq i}, s_j \mid \omega = L).$$

Adding across all such histories then gives

$$\mathbb{P}(\tau_j^i = t \mid \omega = H) \geq \mathbb{P}(\tau_j^i = t \mid \omega = L).$$

□

Proof of Lemma 5. Define $C = A \cap B$ and $D = B \setminus A$. If $\mathbb{P}(s_i \in D \mid \omega = H) = 0$ then the result is immediate, so assume $\mathbb{P}(s_i \in D \mid \omega = H) = 0$. Then

$$\mathbb{P}(s_i \in A \mid s_i \in B,\ \omega = H) = \frac{\mathbb{P}(s_i \in C \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H) + \mathbb{P}(s_i \in D \mid \omega = H)}.$$

Now,

$$\frac{\mathbb{P}(s_i \in D \mid \omega = L)}{\mathbb{P}(s_i \in C \mid \omega = L)} \leq \alpha^2 \frac{\mathbb{P}(s_i \in D \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H)},$$

so

$$\begin{aligned}\frac{\log(\frac{\mathbb{P}(s_i \in C \mid \omega = L)}{\mathbb{P}(s_i \in C \mid \omega = L) + \mathbb{P}(s_i \in D \mid \omega = L)})}{\log(\frac{\mathbb{P}(s_i \in C \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H) + \mathbb{P}(s_i \in D \mid \omega = H)})} &= \frac{\log(1 + \frac{\mathbb{P}(s_i \in D \mid \omega = L)}{\mathbb{P}(s_i \in C \mid \omega = L)})}{\log(1 + \frac{\mathbb{P}(s_i \in D \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H)})} \\ &\leq \frac{\log(1 + \alpha^2 \frac{\mathbb{P}(s_i \in D \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H)})}{\log(1 + \frac{\mathbb{P}(s_i \in D \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H)})}.\end{aligned}$$

Now, for any $x \in (0, 1]$, $\log(1 + \alpha^2 x) \leq \alpha^2 x$ and $\log(1 + x) \geq x/2$, so

$$\frac{\log(1 + \alpha^2 x)}{\log(1 + x)} \leq 2\alpha^2.$$

Since $\mathbb{P}(s_i \in A \mid s_i \in B,\ \omega = H) \geq 1/2$, it follows that $\frac{\mathbb{P}(s_i \in D \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H)} \leq 1$, so

$$\frac{\log(\frac{\mathbb{P}(s_i \in C \mid \omega = L)}{\mathbb{P}(s_i \in C \mid \omega = L) + \mathbb{P}(s_i \in D \mid \omega = L)})}{\log(\frac{\mathbb{P}(s_i \in C \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H) + \mathbb{P}(s_i \in D \mid \omega = H)})} \leq 2\alpha^2$$

and hence

$$\log(\frac{\mathbb{P}(s_i \in C \mid \omega = L)}{\mathbb{P}(s_i \in C \mid \omega = L) + \mathbb{P}(s_i \in D \mid \omega = L)}) \geq 2\alpha^2 \log(\frac{\mathbb{P}(s_i \in C \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H) + \mathbb{P}(s_i \in D \mid \omega = H)}).$$

so

$$\begin{aligned}\mathbb{P}(s_i \in A \mid s_i \in B,\ \omega = L) &= \frac{\mathbb{P}(s_i \in C \mid \omega = L)}{\mathbb{P}(s_i \in C \mid \omega = L) + \mathbb{P}(s_i \in D \mid \omega = L)} \\ &\geq \left(\frac{\mathbb{P}(s_i \in C \mid \omega = H)}{\mathbb{P}(s_i \in C \mid \omega = H) + \mathbb{P}(s_i \in D \mid \omega = H)}\right)^{2\alpha^2} \\ &= \mathbb{P}(s_i \in A \mid s_i \in B,\ \omega = H)^{2\alpha^2}.\end{aligned}$$

□

Proof of Lemma 6. Since the s_i are conditionally independent,

$$\mathbb{P}(A \mid B, \omega = L) = \prod_{i \in S} \mathbb{P}(s_i \in A_i \mid s_i \in B_i, \omega = L)$$

and

$$\mathbb{P}(A \mid B, \omega = H) = \prod_{i \in S} \mathbb{P}(s_i \in A_i \mid s_i \in B_i, \omega = H).$$

Since $\mathbb{P}(A \mid B, \omega = H) \geq \frac{1}{2}$, it follows that $\mathbb{P}(s_i \in A_i \mid B_i, \omega = H) \geq \frac{1}{2}$ for each $i \in S$, and hence by the previous lemma,

$$\mathbb{P}(s_i \in A_i \mid B_i, \omega = L) \geq \mathbb{P}(s_i \in A_i \mid B_i, \omega = H)^{2\alpha^2}$$

for each $i \in S$. Thus,

$$\mathbb{P}(A \mid B, \omega = L) \geq \mathbb{P}(A \mid B, \omega = H)^{2\alpha^2} \geq \left(\frac{1}{2}\right)^{2\alpha^2}.$$

□

Proof of Lemma 7. By no spontaneous investment, $\tau^i_{i+1} = t$ if and only if $\tau^i_{i+1+j} = t - j$ for $j = 0, \dots, t$.

For $j = 0, \dots, t$, let $B_j \subseteq [0, 1]$ be the set of signals such that agent $i + (t + 1) - j$ invests in period 0 if and only if $s_{i+(t+1)-j} \notin B_j$, and for $j = 1, \dots, t$, let $A_j \subseteq [0, 1]$ be the set of signals such that agent $i+(t+1)-j$ invests in period j if $s_{i+(t+1)-j} \in A_j$, agent $i + (t + 1) - j + 1$ invests in period $j - 1$, and agent $i + (t + 1) - j + 1$ does not invest before period j. Let A be the event that $s_{i+(t+1)-j} \in A_j$ for $j = 1, \dots, t$, and let B be the event that $s_{i+(t+1)-j} \in B_j$ for $j = 1, \dots, t$.

Then

$$\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega) = \mathbb{P}(A \mid B, s_{i+(t+1)} \notin B_0, \omega) = \mathbb{P}(A \mid B, \omega).$$

Hence, if $\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = H) \geq \frac{1}{2}$, then $\mathbb{P}(A \mid B, \omega = H) \geq \frac{1}{2}$, so by the lemma,

$$\mathbb{P}(\tau^i_{i+1} = t \mid C_{i,t}, \omega = L) = \mathbb{P}(A \mid B, \omega = L) \geq \left(\frac{1}{2}\right)^{2\alpha^2}.$$

□

Chapter 2

THE SPEED OF SEQUENTIAL ASYMPTOTIC LEARNING

2.1 Introduction

When making decisions, we often rely on the decisions that others before us have made. Sequential learning models have been used to understand different phenomena that occur when many individuals make decisions based on the observed actions of others. These include herd behavior (cf. Banerjee (1992)), where many agents make the same choice, as well as informational cascades (e.g., Bikhchandani et al. (1992)), where the actions of the first few agents provide such compelling evidence that later agents no longer have incentive to consider their own private information.

Such results on how information aggregation can fail are complemented by results which demonstrate that when private signals are arbitrarily strong, learning is robust to this kind of collapse Smith and Sorensen (2000). In particular, in a process called asymptotic learning (see, e.g., Acemoglu et al. (2011)), agents will eventually choose the correct action and their beliefs will converge to the truth. A question that has not been answered in the literature is: how quickly does this happen? And how does the speed of learning compare to a setting in which agents observe signals rather than actions?

We consider the classical setting of a binary state of nature and binary actions, where each of the two actions is optimal at one of the states. The agents receive private signals that are independent conditioned on the state. These signals are unbounded, in the sense that an agent's posterior belief regarding the state can be arbitrarily close to both 0 and 1. The agents are exogenously ordered, and, at each time period, a single agent takes an action, after observing the actions of her predecessors.

We measure the speed of learning by studying how the public belief evolves as more and more agents act. Consider an outside observer who observes the actions of the sequence of agents. The public belief is the posterior belief that such an outside observer assigns to the correct state of nature. It provides a measure of how well the population has learned the state. Since signals are unbounded, the public belief tends to 1 over time Smith and Sorensen (2000); equivalently, the corresponding log-likelihood ratio tends to infinity. As the outside observer may also be interested in learning the state, it is natural to ask how quickly she converges to the correct belief, and, in particular, to understand her asymptotic speed of learning when observing actions. Asymptotic rates of convergence are an important tool in the study of inference processes in statistical theory, and have also been studied in social learning models in the Economics literature (e.g., Duffie and Manso (2007), Duffie et al. (2009), Vives (1993)).

When agents observe the *signals* (rather than actions) of all of their predecessors, this log-likelihood ratio is asymptotically linear. Thus, it cannot grow faster than linearly when the agents observe actions. Our first main finding is that when observing actions, the log-likelihood ratio always grows sub-linearly. Equivalently, the public belief converges sub-exponentially to 1. Our second main finding is that, depending on the choice of private signal distributions, the log-likelihood ratio can grow at a rate that is arbitrarily close to linear.

We next analyze the specific canonical case of Gaussian private signals. Here we calculate precisely the asymptotic behavior of the log-likelihood ratio of the public belief. We show that learning from actions is significantly slower than learning from signals: the log-likelihood ratio behaves asymptotically as $\sqrt{\log t}$. To calculate this we develop a technique that allows, much more generally, for the long-term evolution of the public belief to be calculated for a large class of signal distributions.

Since, in our setting of unbounded signals, agents eventually take the correct action, an additional, natural measure of the speed of learning is the expected time at which this happens: how long does it take until no more mistakes are made? We call this the *time to learn*.

We show that the expected time to learn depends crucially on the signal distributions. For distributions, such as the Gaussian, in which strong signals occur with very small probability, we show that the expected time to learn is infinite.[1] However, when strong signals are less rare, this expectation is finite.[2] Intuitively, when strong signals are rare, agents are more likely to emulate their predecessors, and so it may take a long time for a mistake to be corrected.

Finally, in the Gaussian case, we study another measure of the speed of learning. Namely, we consider directly how the probability of choosing the incorrect action varies as agents see more and more of the other agents' decisions before making their own. We find that this probability is asymptotically no less than $1/t^{1+\varepsilon}$ for any

[1] In the benchmark case of observed signals this time is finite, for any signal distribution.

[2] This result disproves a conjecture of Sorensen Sorensen (1996) (page 36).

$\varepsilon > 0$. In contrast, when agents can observe the private signals of their predecessors, the probability of mistake decays exponentially, and so also in this sense learning from signals is much faster than learning from actions.

Related literature

Several previous studies have considered the same question. Chamley Chamley (2004b) gives an estimate for the evolution of the public belief for a class of private signal distributions with fat tails. He also studies the speed of convergence in the Gaussian case using a computer simulation. Sorensen (Sorensen, 1996, Lemma 1.9) has published a claim related to our Theorem 4, with an unfinished proof. Also in Sorensen (1996), Sorensen shows that the expected time to learn is infinite for some signal distributions, and conjectures that it is always infinite, which we show to not be true. In Smith and Sorensen (1996), an early version of Smith and Sorensen (2000), the question of the time to learn is also addressed, and an example is given in which the time to learn is infinite, but is finite conditioned on one of the states. A concurrent paper by Rosenberg and Vieille Rosenberg and Vieille (2017) studies related questions. In particular they study the time until the first correct action, as well as the number of incorrect actions—which are related to our time to learn—and characterize when they have finite expectations.

A related model is studied by Lobel, Acemoglu, Dahleh and Ozdaglar Lobel et al. (2009), who consider agents who also act sequentially, but do not observe all of their predecessors' actions. They study how the speed of learning varies with the network structure. Vives Vives (1993), in a paper with a very similar spirit to ours, studies the speed of sequential learning in a model with actions chosen from a continuum, and where agents observe a noisy signal about their predecessors' actions. He similarly shows that learning is significantly slower than in the benchmark case. An overview of this literature is given by Vives in his book (Vives, 2010, Chapter 6).

2.2 Model

Let $\theta \in \{-1, +1\}$ be the true state of the world, with each state a priori equally likely[3]. Each rational agent $t \in \{1, 2, \ldots\}$ receives a private signal s_t. The signals are i.i.d. conditioned on θ: if $\theta = +1$ they have cumulative distribution function (CDF) F_+ and if $\theta = -1$ they have CDF F_-.[4] We assume that F_+ and F_- are absolutely continuous with respect to each other, so that private signals never completely reveal the state.

Let

$$L_t = \log \frac{\mathbb{P}(\theta = +1|s_t)}{\mathbb{P}(\theta = -1|s_t)}$$

be the log-likelihood ratio of the belief induced by the agent's private signal. We assume that private signals are unbounded, in the sense that L_t is unbounded: for every $M \in \mathbb{R}$ the probability that $L_t > M$ is positive, as is the probability that $L_t < -M$. We denote by G_+ and G_- the conditional CDFs of L_t.

The agents act sequentially, with agent t acting after observing the actions of agents $\{1, \ldots, t-1\}$. The utility of the action $a_t \in \{-1, +1\}$ is 1 if $a_t = \theta$ and 0 otherwise.

Denote the public belief by

$$\mu_t = \mathbb{P}(\theta = +1|a_1, \ldots, a_{t-1}).$$

This is the posterior held by an outside observer after recording the actions of the first $t-1$ agents. We denote by ℓ_t the log-likelihood ratio of the public belief:

$$\ell_t = \log \frac{\mu_t}{1 - \mu_t}.$$

In equilibrium, agent t chooses $a_t = +1$ iff[5]

$$\log \frac{\mathbb{P}(\theta = +1|a_1, \ldots, a_{t-1}, s_t)}{\mathbb{P}(\theta = -1|a_1, \ldots, a_{t-1}, s_t)} > 0.$$

A simple calculation shows that this occurs iff

$$\ell_t + L_t > 0.$$

Now, another straightforward calculation shows that when $a_t = +1$,

$$\ell_{t+1} = \ell_t + D_+(\ell_t), \tag{2.1}$$

[3]We make this simplification of a (1/2,1/2) prior to reduce the complexity of the presentation, but all results hold for general priors.

[4]One could consider signals that take values in a general measurable space (rather than $\mathbb{R}$), but the choice of $\mathbb{R}$ is in fact without loss of generality, since all standard measurable spaces are isomorphic.

[5]For simplicity, we assume that agents choose action -1 when indifferent. This will have no impact on our results.

where

$$D_+(x) = \log \frac{1 - G_+(-x)}{1 - G_-(-x)}.$$

Likewise, when $a_t = -1$,

$$\ell_{t+1} = \ell_t + D_-(\ell_t),$$

where

$$D_-(x) = \log \frac{G_+(-x)}{G_-(-x)}.$$

We can interpret $D_+(\ell_t)$ and $D_-(\ell_t)$ as the contributions of agent t's action to the public belief.

2.3 The evolution of public belief

Consider a baseline model, in which each agent observes the private signals of all of her predecessors. In this case the public log-likelihood ratio $\tilde{\ell}_t$ would equal the sum

$$\tilde{\ell}_t = \sum_{\tau=1}^{t} L_\tau.$$

Conditioned on the state this is the sum of i.i.d. random variables, and so by the law of large numbers we have that the limit $\lim_t \tilde{\ell}_t/t$ would—conditioned on (say) $\theta = +1$—equal the conditional expectation of L_t, which is positive.[6]

Sub-linear public beliefs

Our first main result shows that when agents observe actions rather than signals, the public log-likelihood ratio grows sub-linearly, and so learning from actions is always slower than learning from signals.

Theorem 4. *It holds with probability 1 that* $\lim_t \ell_t/t = 0$.

Our second main result shows that, depending on the choice of private signal distributions, ℓ_t can grow at a rate that is arbitrarily close to linear: given any sub-linear function r_t, it is possible to find private signal distributions so that ℓ_t grows as fast as r_t.

[6] In fact, $\mathbb{E}(L_t|\theta = +1)$ is equal to the Kullback-Leibler divergence between F_+ and F_-, which is positive as long as the two distributions are different.

Theorem 5. *For any* $r\colon \mathbb{N} \to \mathbb{R}_{>0}$ *such that* $\lim_t r_t/t = 0$ *there exists a choice of CDFs* F_- *and* F_+ *such that*

$$\liminf_{t\to\infty} \frac{|\ell_t|}{r_t} > 0$$

with probability 1.

For example, for some choice of private signal distributions, ℓ_t grows asymptotically at least as fast as $t/\log t$, which is sub-linear but (perhaps) close to linear.

Long-term behavior of public beliefs

We next turn to estimating more precisely the long-term behavior of the public log-likelihood ratio ℓ_t. Since signals are unbounded, agents learn the state, so that ℓ_t tends to $+\infty$ if $\theta = +1$, and to $-\infty$ if $\theta = -1$. In particular ℓ_t stops changing sign from some t on, with probability 1; all later agents choose the correct action.

We consider without loss of generality the case that $\theta = +1$, so that ℓ_t is positive from some t on. Thus, recalling (2.1), we have that from some t on,

$$\ell_{t+1} = \ell_t + D_+(\ell_t).$$

This is the recurrence relation that we need to solve in order to understand the long term evolution of ℓ_t. To this end, we consider the corresponding differential equation:

$$\frac{\mathrm{d}f}{\mathrm{d}t}(t) = D_+(f(t)).$$

Recall that G_- is the CDF of the private log-likelihood ratio L_t, conditioned on $\theta = -1$. We show (Lemma 8) that $D_+(x)$ is well approximated by $G_-(-x)$ for high x, in the sense that

$$\lim_{x\to\infty} \frac{D_+(x)}{G_-(-x)} = 1.$$

In some applications (including the Gaussian one, which we consider below), the expression for G_- is simpler than that for D_+, and so one can instead consider the differential equation

$$\frac{\mathrm{d}f}{\mathrm{d}t}(t) = G_-(-f(t)). \tag{2.2}$$

This equation can be solved analytically in many cases in which G_- has a simple form. For example, if $G_-(-x) = \mathrm{e}^{-x}$ then $f(t) = \log(t + c)$, and if $G_-(-x) = x^{-k}$ then $f(t) = ((k+1)\cdot t + c)^{1/(k+1)}$.

We show that solutions to this equation have the same long term behavior as ℓ_t, given that G_- satisfies some regularity conditions.

Theorem 6. *Suppose that G_- and G_+ are continuous, and that the left tail of G_- is convex and differentiable. Suppose also that $f \colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfies*

$$\frac{\mathrm{d}f}{\mathrm{d}t}(t) = G_-(-f(t)) \tag{2.3}$$

for all sufficiently large t. Then conditional on $\theta = +1$,

$$\lim_{t\to\infty} \frac{\ell_t}{f(t)} = 1$$

with probability 1.

The condition[7] on G_- is satisfied when the random variables L_t (i.e., the log-likelihood ratios associated with the private signals), conditioned on $\theta = -1$, have a distribution with a probability density function that is monotone decreasing for all x less than some x_0. This is the case for the normal distribution, and for practically every non-atomic distribution one may encounter in the standard probability and statistics literatures.

Gaussian signals

In the Gaussian case, F_+ is Normal with mean +1 and variance σ^2, and F_- is Normal with mean -1 and the same variance. A simple calculation shows that G_- is the Gaussian cumulative distribution function, and so we cannot solve the differential equation (2.2) analytically. However, we can bound $G_-(x)$ from above and from below by functions of the form $\mathrm{e}^{-c\cdot x^2}/x$. For these functions the solution to (2.2) is of the form $f(t) = \sqrt{\log t}$, and so we can use Theorem 6 to deduce the following.

Theorem 7. *When private signals are Gaussian, then conditioned on $\theta = +1$,*

$$\lim_{t\to\infty} \frac{\ell_t}{(2\sqrt{2}/\sigma)\cdot\sqrt{\log t}} = 1$$

with probability 1.

Recall, that when private signals are observed, the public log-likelihood ratio ℓ_t is asymptotically *linear*. Thus, learning from actions is far slower than learning from signals in the Gaussian case.

[7] By "the left tail of G_- is convex and differentiable" we mean that there is some x_0 such that, restricted to $(-\infty, x_0)$, G_- is convex and differentiable.

The expected time to learn

When private signals are unbounded then with probability 1 the agents eventually all choose the correct action $a_t = \theta$. A natural question is: how long does it take for that to happen? Formally, we define the *time to learn*

$$T_L = \min\{t \,:\, a_\tau = \theta \text{ for all } \tau \geq t\},$$

and study its expectation. Note that in the baseline case of observed signals T_L has finite expectation, since the probability of a mistake at time t decays exponentially with t.

We first study the expectation of T_L in the case of Gaussian signals. To this end we define the *time of first mistake* by

$$T_1 = \min\{t \,:\, a_t \neq \theta\}$$

if $a_t \neq \theta$ for some t, and by $T_1 = 0$ otherwise. We calculate a lower bound for the distribution of T_1, showing that it decays at most as fast as $1/t$.

Theorem 8. *When private signals are Gaussian then for every $\varepsilon > 0$ there exists a $k > 0$ such that for all t*

$$\mathbb{P}(T_1 = t) \geq \frac{k}{t^{1+\varepsilon}}.$$

Thus T_1 has a very thick tail, decaying far slower than the exponential decay of the baseline case. In particular, T_1 has infinite expectation, and so, since $T_L > T_1$, the expectation of the time to learn T_L is also infinite.

In contrast, we show that when private signals have thick tails—that is, when the probability of a strong signal vanishes slowly enough—then the time to learn has finite expectation. In particular, we show this when the left tail of G_- and the right tail of G_+ are polynomial.[8]

Theorem 9. *Assume that $G_-(-x) = c \cdot x^{-k}$ and that $G_+(x) = 1 - c \cdot x^{-k}$ for some $c > 0$ and $k > 0$, and for all x greater than some x_0. Then $\mathbb{E}(T_L) < \infty$.*

[8] Recall that G_- is the conditional cumulative distribution function of the private log-likelihood ratios L_t.

An example of private signal distributions F_+ and F_- for which G_- and G_+ have this form is given by the probability density functions

$$f_-(x) = \begin{cases} c \cdot \mathrm{e}^{-x} x^{-k-1} & \text{when } 1 \le x \\ 0 & \text{when } -1 < x < 1 \\ c \cdot (-x)^{-k-1} & \text{when } x \le -1. \end{cases}$$

and $f_+(x) = f_-(-x)$, for an appropriate choice of normalizing constant $c > 0$. In this case $G_-(-x) = 1 - G_+(x) = \frac{c}{k} x^{-k}$ for all $x > 1$.[9]

The proof of Theorem 9 is rather technically involved, and we provide here a rough sketch of the ideas behind it.

We say that there is an *upset* at time t if $a_{t-1} \neq a_t$. We denote by Ξ the random variable which assigns to each outcome the total number of upsets

$$\Xi = |\{t \,:\, a_{t-1} \neq a_t\}|.$$

We say that there is a *run* of length m from time t if $a_t = a_{t+1} = \cdots = a_{t+m-1}$. As we will condition on $\theta = +1$ in our analysis, we say that a run from time t is *good* if $a_t = 1$ and *bad* otherwise. A trivial but important observation is that the number of maximal finite runs is equal to the number of upsets, and so, if $\Xi = n$, and if $T_L = t$, then there is at least one run of length t/n before time t. Qualitatively, this implies that if the number of upsets is small, and if the time to learn is large, then there is at least one long run before the time to learn.

We show that it is indeed unlikely that Ξ is large: the distribution of Ξ has an exponential tail. Incidentally, this holds for *any* private signal distribution:

Proposition 8. *For every private signal distribution there exist $c > 0$ and $0 < \gamma < 1$ such that for all $n > 0$*

$$\mathbb{P}(\Xi \ge n) \le c\gamma^n.$$

Intuitively, this holds because whenever an agent takes the correct action, there is a non-vanishing probability that all subsequent agents will also do so, and no more upsets will occur.

[9] Theorem 9 can be proved for other thick-tailed private signal distributions: for example, one could take different values of c and k for G_- and G_+, or one could replace their thick polynomial tails by even thicker logarithmic tails. For the sake of readability we choose to focus on this case.

Thus, it is very unlikely that the number of upsets Ξ is large. As we observe above, when Ξ is small then the time to learn T_L can only be large if at least one of the runs is long. When G_- has a thin tail then this is possible; indeed, Theorem 8 shows that the first finite run has infinite expected length when private signals are Gaussian. However, when G_- has a thick, polynomial left tail of order x^{-k}, we show that it is very unlikely for any run to be long: the probability that there is a run of length n decays at least as fast as $\exp(-n^{k/(k+1)})$, and in particular runs have finite expected length. Intuitively, when strong signals are rare then runs tend to be long, as agents are likely to emulate their predecessor. Conversely, when strong signals are more likely then agents are more likely to break a run, and so runs tend to be shorter.

Putting together these insights, we conclude that it is unlikely that there are many runs, and, in the polynomial signal case, it is unlikely that runs are long. Thus T_L has finite expectation.

Probability of taking the wrong action

Yet another natural metric of the speed of learning is the probability of mistake

$$p_t = \mathbb{P}(a_t \neq \theta).$$

Calculating the asymptotic behavior of p_t seems harder to tackle.

For the Gaussian case, while we cannot estimate p_t precisely, Theorem 8 immediately implies a lower bound: p_t is at least $k/t^{1+\varepsilon}$, for every $\varepsilon > 0$ and k that depends on ε. This is much larger than the exponentially vanishing probability of mistake in the revealed signal baseline case.

More generally, we can use Theorem 4 to show that p_t vanishes sub-exponentially for any signal distribution, in the sense that

$$\lim_{t\to\infty} \frac{1}{t} \log p_t = 0.$$

To see this, note that the probability of mistake at time $t-1$, conditioned on the observed actions, is exactly equal to

$$\min\{\mu_t, 1-\mu_t\};$$

where we recall that

$$\mu_t = \mathbb{P}(\theta = +1 | a_1, \ldots, a_{t-1}) = \frac{e^{\ell_t}}{e^{\ell_t}+1}$$

is the public belief. This is due to the fact that if the outside observer, who holds belief μ_t, had to choose an action, she would choose a_{t-1}, the action of the last player she observed, a player who has strictly more information than her. Thus

$$p_t = \mathbb{E}(\min\{\mu_t, 1 - \mu_t\}) = \mathbb{E}\left(\frac{1}{\mathrm{e}^{|\ell_t|} + 1}\right),$$

and since, by Theorem 4, $|\ell_t|$ is sub-linear, it follows that p_t is sub-exponential.

2.4 Conclusion

In this paper we consider a classical setting of sequential asymptotic learning from actions of others. We show that learning from actions is slow, as compared to the speed of learning when observing others' private signals, in the sense that the public log-likelihood ratio tends more slowly to infinity. However, it is possible to approach the linear rate of learning from signals and achieve any sub-linear rate.

We calculate the speed of learning precisely in the case of Normal private signals (among a large class of private signal distributions) and show that learning is very slow. We also show that in the Gaussian case the expected time to learn is infinite, as opposed to cases of more thick-tailed distributions, in which it is finite.

For the Gaussian case we also provide a lower bound for the probability of mistake. Finding a matching upper bound seems beyond our reach at the moment, and provides a compelling open problem for further research.

2.5 Sub-linear learning

Before proving our main theorems we make the observation (which has appeared before, e.g., Chamley (2004b)) that *the log-likelihood ratio of the log-likelihood ratio is the log-likelihood ratio*. Formally, if ν_+ and ν_- are the conditional distributions of the private log-likelihood ratio L_t (i.e., have CDFs G_+ and G_-), then

$$\log \frac{\mathrm{d}\nu_+}{\mathrm{d}\nu_-}(x) = x.$$

It follows that

$$G_+(x) = \int_{-\infty}^{x} \mathrm{d}\nu_+(\zeta) = \int_{-\infty}^{x} \mathrm{e}^{\zeta}\, \mathrm{d}\nu_-(\zeta). \tag{2.4}$$

Our first lemma shows that asymptotically, D_+ behaves like the left tail of G_-, and D_- behaves like the right tail of G_+.

Lemma 8.

$$\lim_{x\to\infty} \frac{D_+(x)}{G_-(-x)} = 1 \textit{ and } \lim_{x\to-\infty} \frac{D_-(x)}{G_+(-x)-1} = 1.$$

Proof. By definition,

$$D_+(x) = \log \frac{1-G_+(-x)}{1-G_-(-x)}.$$

Since $\log(1-z) = -z + O(z^2)$, it holds for all x large enough that

$$D_+(x) > G_-(-x) - 2 \cdot G_+(-x).$$

Applying (2.4) yields

$$D_+(x) > \int_{-\infty}^{-x} (1 - 2\mathrm{e}^{\zeta})\,\mathrm{d}\nu_-(\zeta),$$

and so for any ε and all x large enough,

$$D_+(x) > (1-\epsilon) \cdot \int_{-\infty}^{-x} \mathrm{d}\nu_-(\zeta) = (1-\epsilon)G_-(-x).$$

Using the same approximation of the logarithm, we have that

$$D_+(x) < (1+\epsilon)G_-(-x) - G_+(-x) < (1+\epsilon)G_-(-x).$$

The statement for D_+ now follows by taking ε to zero. The corresponding bounds on D_- follow by identical arguments. □

Proof of Theorem 4. Condition on $\theta = +1$. Then ℓ_t is with probability 1 positive from some point on, and all agents take action +1 from this point on. Hence, for all t large enough,

$$\ell_{t+1} = \ell_t + D_+(\ell_t).$$

By Lemma 8, we know that $\lim_x D_+(x) = 0$. Hence for every $\epsilon > 0$ and all t large enough, $|\ell_{t+1} - \ell_t| < \epsilon$. It follows that the limit $\lim_t \ell_t/t = 0$. The analysis of the case $\theta = -1$ is identical. □

Proof of Theorem 5. Given r_t, we will construct private signal distributions such that $\liminf_t |\ell_t|/r_t > 0$ with probability one. These distributions will furthermore have the property that $D_+(x) = -D_-(-x)$. As a consequence we have that regardless of the action chosen by the agent, as long as the sign of the action is equal to that of ℓ_t (which happens from some point on w.p. 1),

$$|\ell_{t+1}| = |\ell_t| + D_+(|\ell_t|).$$

Intuitively, if we can choose private signal distributions that make $D_+(x)$ decay very slowly, then ℓ_t will be very close to being linear.

Formally, and by elementary considerations, the theorem will follow if, for every $Q\colon \mathbb{R} \to \mathbb{R}_{>0}$ with $\lim_{x\to\infty} Q(x) = 0$, we can find CDFs such that $D_+(x) = -D_-(-x)$ and $\liminf_{x\to\infty} D_+(x)/Q(x) > 0$.

Fix any Q such that $\lim_{x\to\infty} Q(x) = 0$, but assume without loss of generality that $Q(x)$ is monotone decreasing.[10] Define a finite measure ν on the integers by

$$\nu(n) = \frac{Q(n-1) - Q(n)}{\mathrm{e}^n}$$

and

$$\nu(-n) = Q(n-1) - Q(n)$$

for all $n \geq 0$. Note that ν is indeed finite since

$$C := \sum_{n=-\infty}^{\infty} \nu(n) \leq 2Q(-1).$$

Note also that

$$\sum_{n=-\infty}^{\infty} \nu(n) \cdot \mathrm{e}^n$$

is likewise equal to C.

Let the private signal distributions be given by

$$\mathbb{P}(s_t = n | \theta = +1) = C^{-1}\nu(n)\mathrm{e}^n$$

and

$$\mathbb{P}(s_t = n | \theta = -1) = C^{-1}\nu(n).$$

Then

$$L_t = \log \frac{\mathbb{P}(s_t|\theta = +1)}{\mathbb{P}(s_t|\theta = -1)} = s_t,$$

[10]If Q is not monotone decreasing then consider instead $Q'(x) = \sup_{y \geq x} Q(y)$.

the distribution of L_t is identical to that of s_t, and so $G_+ = F_+$ and $G_- = F_-$. By our definition of F_-, we have that for $x > 0$

$$G_-(-x) = C^{-1} \cdot Q(\lceil x \rceil - 1). \tag{2.5}$$

Now, by Lemma 8, we know that

$$(1 - \epsilon) \cdot G_-(-x) < D_+(x) < (1 + \epsilon) \cdot G_-(-x),$$

for any $\epsilon > 0$ and all x large enough. It follows that

$$\liminf_{x \to \infty} \frac{D_+(x)}{Q(x)} = \liminf_{x \to \infty} \frac{G_-(-x)}{Q(x)},$$

which, by (2.5) equals

$$\liminf_{x \to \infty} \frac{C^{-1} Q(\lceil x \rceil - 1)}{Q(x)} \geq C^{-1}.$$

□

2.6 Long-term behavior of public belief

The primary goal of this section is to prove Theorem 6, which states that public belief is asymptotically given by the solution to the differential equation (2.3). The proof of this theorem uses two general lemmas regarding recurrence relations. We state these lemmas now and prove them later. The first lemma states that two similar recurrence relations yield similar solutions. The second shows that the solution to a recurrence relation (of the type we are interested in) is well approximated by the solution to the corresponding differential equation.

Lemma 9. *Let $A, B \colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ be continuous, eventually monotone decreasing, and tending to zero.*

Let (a_t) and (b_t) be sequences satisfying the recurrence relations

$$a_{t+1} = a_t + A(a_t)$$
$$b_{t+1} = b_t + B(b_t).$$

Suppose

$$\lim_{x \to \infty} \frac{A(x)}{B(x)} = 1.$$

Then

$$\lim_{t \to \infty} \frac{a_t}{b_t} = 1.$$

Lemma 10. *Assume that $A\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a continuous function with a convex differentiable tail, and that $A(x)$ goes to 0 as x goes to ∞. Let (a_t) be any sequence satisfying the recurrence equation $a_{t+1} = a_t + A(a_t)$, and suppose there is a function $f : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ with $f'(t) = A(f(t))$ for all sufficiently large t. Then*

$$\lim_{t\to\infty} \frac{f(t)}{a_t} = 1.$$

Given these lemmas, we are ready to prove our theorem.

Proof of Theorem 6. Let (a_t) be any sequence in $\mathbb{R}_{>0}$ satisfying:

$$a_{t+1} = a_t + G_-(-a_t).$$

Then by Lemma 10, the sequence (a_t) is well approximated by $f(t)$, the solution to the corresponding differential equation:

$$\lim_{t\to\infty} \frac{a_t}{f(t)} = 1.$$

Now, conditional on $\theta = +1$, all agents take action $+1$ from some point on with probability 1. Thus, with probability 1,

$$\ell_{t+1} = \ell_t + D_+(\ell_t)$$

for all sufficiently large t. Further, by Lemma 8,

$$\lim_{x\to\infty} \frac{D_+(x)}{G_-(-x)} = 1.$$

So by Lemma 9,

$$\lim_{t\to\infty} \frac{\ell_t}{a_t} = 1$$

with probability 1. Thus, we have

$$\lim_{t\to\infty} \frac{\ell_t}{f(t)} = \lim_{t\to\infty} \frac{\ell_t}{a_t} \cdot \frac{a_t}{f(t)} = 1$$

with probability 1. □

Proofs of Lemmas 9 and 10

Proof of Lemma 9. We prove the claim in two steps. First, we show that for every $\varepsilon > 0$ there are infinitely many times t such that

$$(1-\varepsilon)a_t \leq b_t \leq (1+\varepsilon)a_t. \tag{2.6}$$

Second, we show that if (2.6) holds for some t large enough, then it holds for all $t' > t$, proving the claim.

We start with step 1. Assume without loss of generality that $a_t \leq b_t$ for infinitely many values of t. Fix $\varepsilon > 0$. To show that $(1-\varepsilon)a_t \leq b_t \leq (1+\varepsilon)a_t$ holds for infinitely many values of t, let $x_0 > 1$ be such that for all $x > x_0$ it holds that A and B are monotone decreasing,

$$A(x), B(x) < \varepsilon < 1$$

and

$$(1-\varepsilon/2)A(x) < B(x) < (1+\varepsilon/2)A(x). \tag{2.7}$$

Assume that $a_t, b_t > x_0$; this will indeed be the case for t large enough, since A and B are positive and continuous, and so both a_t and b_t are monotone increasing and tend to infinity. So

$$B(b_t) < (1+\varepsilon/2)A(b_t) \leq (1+\varepsilon/2)A(a_t),$$

where the first inequality follows from (2.7), and the second follows from the fact that A is monotone decreasing and $a_t < b_t$. Since $B(b(t)) = b_{t+1} - b(t)$ and $A(a_t) = a_{t+1} - a(t)$ we have shown that

$$b_{t+1} - b_t < (1+\varepsilon/2)(a_{t+1} - a_t),$$

and so eventually $b_t \leq (1+\varepsilon)a_t$. Also, notice that the first time this obtains, we also have that the left inequality in (2.6) holds at the same moment:

$$b_t > b_{t-1} > a_{t-1} = a_t - (a_t - a_{t-1}) > a_t - \varepsilon > a_t - \varepsilon a_t = (1-\varepsilon)a_t.$$

This completes the first step. Now we go to step 2. Here we show that if (2.6) holds for large enough t then it holds for all $t' > t$.

Fix $\varepsilon > 0$, and let x_0 be defined as above. Suppose that $(1-\varepsilon)a_t < b_t < (1+\varepsilon)a_t$, with $a_t, b_t > x_0$. Assume without loss of generality that $b_t \geq a_t$. Then our assumptions and (2.7) imply

$$\begin{aligned} b_{t+1} &= b_t + B(b_t) \\ &< (1+\varepsilon)a_t + (1+\varepsilon)A(b_t). \end{aligned}$$

Because $a_t \leq b_t$ and A is decreasing we have

$$\begin{aligned} b_{t+1} &< (1+\varepsilon)a_t + (1+\varepsilon)A(a_t) \\ &= (1+\varepsilon)a_{t+1}. \end{aligned}$$

For the other direction, note first that

$$b_{t+1} > b_t \geq a_t,$$

by assumption. We can write $a_t = (1-\varepsilon)a_t + \varepsilon a_t$, and since $a_t > x_0 > 1$, $\varepsilon a_t > (1-\varepsilon)\varepsilon$, and so

$$b_{t+1} > (1-\varepsilon)a_t + (1-\varepsilon)\varepsilon.$$

Now, $\varepsilon > A(a_t)$ since $a_t > x_0$, and so

$$\begin{aligned} b_{t+1} &> (1-\varepsilon)a_t + (1-\varepsilon)A(a_t) \\ &= (1-\varepsilon)a_{t+1}. \end{aligned}$$

Thus

$$(1-\varepsilon)a_{t+1} < b_{t+1} < (1+\varepsilon)a_{t+1}, \tag{2.8}$$

as required.

□

Proof of Lemma 10. We restrict the domain of f to the interval (t_0, ∞) such that for $t > t_0$ it already holds that $f'(t) = A(f(t))$. Since A is continuous, $\lim_{t\to\infty} f(t) = \infty$, and so we can also assume that in the interval $(f(t_0), \infty)$ it holds that A is convex and differentiable.

Since f is strictly increasing in (t_0, ∞), it has an inverse f^{-1}. For x large enough define $B(x) = f(f^{-1}(x) + 1) - x$.

Now, let (b_t) be any sequence satisfying the recurrence relation

$$b_{t+1} = b_t + B(b_t).$$

In order to apply Lemma 9, we will first show that

$$\lim_{x\to\infty} \frac{B(x)}{A(x)} = 1.$$

Let $t = f^{-1}(x)$. Such a t exists and is unique for all sufficiently large x, because f is monotone. Notice that by the definitions of $B(x)$ and $f'(x)$

$$\begin{aligned} B(x) &= f(f^{-1}(x) + 1) - x \\ &= f(f^{-1}(x) + 1) - x - f'(f^{-1}(x)) + f'(f^{-1}(x)) \\ &= f(t+1) - f(t) - f'(t) + A(f(t)), \end{aligned}$$

where in the last equality we substitute $t = f^{-1}(x)$. Because f' is positive and decreasing (f is concave) then $f(t+1) - f(t) \geq f'(t+1)$, and so

$$B(x) \geq f'(t+1) - f'(t) + A(f(t)).$$

By the definition of f, $f'(t) = A(f(t))$, and so

$$B(x) \geq A(f(t+1)) - A(f(t)) + A(f(t)) = A(f(t+1)).$$

Again, due to concavity of f we have $f(t+1) \leq f(t) + f'(t)$ and as A is decreasing and convex we get

$$\begin{aligned} B(x) &\geq A(f(t) + f'(t)) \\ &\geq A'(f(t))f'(t) + A(f(t)) \\ &= A'(f(t))A(f(t)) + A(f(t)). \end{aligned}$$

We now substitute back $x = f(t)$:

$$\begin{aligned} B(x) &\geq A'(x)A(x) + A(x) \\ &= A(x)(A'(x) + 1) \end{aligned}$$

so in particular, since $A'(x) \to 0$ as $x \to \infty$,

$$\liminf_{x\to\infty} \frac{B(x)}{A(x)} \geq 1.$$

Now we are going to show that $\limsup_{x\to\infty} \frac{B(x)}{A(x)} \leq 1$ which will conclude the proof. By the definitions of $f^{-1}(x)$ and $B(x)$

$$B(x) = B(f(t)) = f(t+1) - f(t) = \int_t^{t+1} f'(\zeta)\,\mathrm{d}\zeta.$$

As f' is decreasing it follows that

$$B(x) \leq \int_t^{t+1} f'(t)\,\mathrm{d}\zeta = f'(t) = A(f(t)) = A(x).$$

Therefore,

$$\limsup_{x\to\infty} \frac{B(x)}{A(x)} \leq 1.$$

Hence, from these two inequalities we get that

$$\lim_{x\to\infty} \frac{B(x)}{A(x)} = 1.$$

Now notice that, by construction, $f(t+1) = f(t) + B(f(t))$. Thus, by Lemma 9,

$$\lim_{n\to\infty} \frac{f(t)}{a_t} = 1.$$

□

Monotonicity of solutions to a differential equation

We now prove a general lemma regarding differential equations of the form $a'(t) = A(a(t))$. It shows that the solutions to this equation are monotone in A. This is useful for calculating approximate analytic solutions whenever it is impossible to find analytic exact solutions, as is the case of Gaussian signals, in which we use this lemma.

Lemma 11. *Let $A, B\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ be continuous, and let $a, b\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfy $a'(t) = A(a(t))$ and $b'(t) = B(b(t))$ for all sufficiently large t.*

Suppose that

$$\liminf_{x\to\infty} \frac{A(x)}{B(x)} > 1.$$

Then $a(t) > b(t)$ for all sufficiently large t.

Proof. Notice that $a(t)$ and $b(t)$ are eventually monotone increasing and tend to infinity as t tends to infinity. Thus for all x greater than some $x_0 > 0$ large enough, a and b have inverses that satisfy the following differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}x}a^{-1}(x) = \frac{1}{A(x)}$$
$$\frac{\mathrm{d}}{\mathrm{d}x}b^{-1}(x) = \frac{1}{B(x)}.$$

Since $\liminf_x A(x)/B(x) > 1$, we can furthermore choose x_0 so that for all $x \geq x_0$, $A(x) > (1+\varepsilon)B(x)$ for some $\varepsilon > 0$. Thus, for $x > x_0$

$$a^{-1}(x) = a^{-1}(x_0) + \int_{x_0}^{x} \frac{1}{A(x)}\,\mathrm{d}x$$
$$b^{-1}(x) = b^{-1}(x_0) + \int_{x_0}^{x} \frac{1}{B(x)}\,\mathrm{d}x$$

and so

$$\begin{aligned} a^{-1}(x) &< a^{-1}(x_0) + \frac{1}{1+\varepsilon}\int_{x_0}^{x} \frac{1}{B(x)}\,\mathrm{d}x \\ &= a^{-1}(x_0) + \frac{1}{1+\varepsilon}(b^{-1}(x) - b^{-1}(x_0)) \end{aligned}$$

and thus

$$a^{-1}(x) - b^{-1}(x) < -\frac{\varepsilon}{1+\varepsilon}b^{-1}(x) + \left[a^{-1}(x_0) - \frac{1}{1+\varepsilon}b^{-1}(x_0)\right].$$

Since $b^{-1}(x)$ tends to infinity as x tends to infinity, it follows that for all sufficiently large x, $a^{-1}(x) < b^{-1}(x)$. Thus, for all sufficiently large t

$$t = a^{-1}(a(t)) < b^{-1}(a(t)),$$

and so, since $b(t)$ is monotone increasing,

$$b(t) < a(t).$$

□

Eventual monotonicity of public belief update

We end this section with a lemma that shows that under some technical conditions on the left tail of G_-, the function $u_+(x) = x + D_+(x)$ (i.e., the function that determines how the public log-likelihood ratio is updated when the action +1 is taken) is eventually monotone increasing.

Lemma 12. *Suppose G_- has a convex and differentiable left tail. Then the map $u_+(x) = x + D_+(x)$ is monotone increasing for all sufficiently large x.*

Proof. Recall that

$$D_+(x) = \log \frac{1 - G_+(-x)}{1 - G_-(-x)}.$$

Since G_- has a differentiable left tail, it has a derivative $g_-(-x)$ for all x large enough. It then follows from (2.4) that G_+ also has a derivative in this domain, and

$$\begin{aligned} u'_+(x) &= 1 + \frac{g_+(-x)}{1 - G_+(-x)} - \frac{g_-(-x)}{1 - G_-(-x)} \\ &= 1 + \frac{\mathrm{e}^{-x} g_-(-x)}{1 - G_+(-x)} - \frac{g_-(-x)}{1 - G_-(-x)}. \end{aligned}$$

Since $1 - G_-(-x)$ and $1 - G_+(-x)$ tend to 1 as x tends to infinity,

$$\lim_{x \to \infty} u'_+(x) = \lim_{x \to \infty} 1 + \mathrm{e}^{-x} g_-(-x) - g_-(-x).$$

Since G_- is eventually convex, $g_-(-x)$ tends to zero, and therefore

$$\lim_{x \to \infty} u'_+(x) = 1.$$

In particular, $u'_+(x)$ is positive for x large enough, and hence $u_+(x)$ is eventually monotone increasing. □

2.7 Gaussian private signals

Preliminaries

We say that private signals are Gaussian when F_- is the normal distribution with mean -1 and variance σ^2, and F_+ is the normal distribution with mean $+1$ and variance σ^2. To calculate the evolution of ℓ_t, we need to calculate G_+ and G_-, the conditional distributions of the private log-likelihood ratio L_t. Notice that in this case

$$L_t = \log \frac{\mathrm{e}^{-(s_t - 1)^2/2\sigma^2}}{\mathrm{e}^{-(s_t - (-1))^2/2\sigma^2}} = 2 s_t / \sigma^2,$$

so that L_t is simply proportional to the signal s_t. It follows that L_t is also normally distributed, conditioned on the state θ, and that G_+ and G_- are cumulative distribution functions of Gaussians, with variance $4/\sigma^2$.

Notation

In this section and those that follow, we denote by ℓ_t^* the public log-likelihood ratio when all agents before agent t take the correct action. Formally,

$$\ell_t^* = \log \frac{\mathbb{P}(\theta = +1 \mid a_1 = \cdots = a_{t-1} = +1)}{\mathbb{P}(\theta = -1 \mid a_1 = \cdots = a_{t-1} = +1)}.$$

For convenience, we will also use the notation $\mathbb{P}_+(\cdot)$ as shorthand for $\mathbb{P}(\cdot \mid \theta = +1)$.

The evolution of public belief

Proof of Theorem 7. Let $f: \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ be any function such that $f'(t) = G_-(-f(t))$ for all sufficiently large t. Then by Theorem 6,

$$\lim_{t\to\infty} \frac{\ell_t}{f(t)} = 1$$

with probability 1.

Recall from above that L_t is distributed normally, and $G_-(-x)$ is the CDF of a normal distribution with variance $\tau^2 = 4/\sigma^2$.

For $1 > \eta \geq 0$, define

$$F_\eta(x) = \frac{\mathrm{e}^{-\frac{1-\eta}{2\tau^2}x^2}}{x}$$
$$f_\eta(t) = \frac{\sqrt{2}\tau}{\sqrt{1-\eta}}\sqrt{\log(t) + \log\frac{(1-\eta)^2}{2\tau^2}}.$$

By a routine application of L'Hospital's rule, F_0 and F_η are lower and upper bounds for G_-, in the sense that

$$\lim_{x\to\infty} \frac{G_-(-x)}{F_0(x)} = \infty$$
$$\lim_{x\to\infty} \frac{F_\eta(x)}{G_-(-x)} = \infty, \eta > 0.$$

Since $f_\eta'(t) = F_\eta(f_\eta(t))$ for all sufficiently large t, we have by Lemma 11 that for any $\eta > 0$,

$$f_0(t) < f(t) < f_\eta(t)$$

for all sufficiently large t. So

$$\liminf_{t\to\infty} \frac{f(t)}{\sqrt{2}\tau\sqrt{\log t}} = \liminf_{t\to\infty} \frac{f(t)}{f_0(t)} \geq 1$$

and for any $\eta > 0$,

$$\limsup_{t\to\infty} \frac{f(t)}{\sqrt{2}\tau\sqrt{\log t}} = \frac{1}{\sqrt{1-\eta}} \cdot \limsup_{t\to\infty} \frac{f(t)}{f_\eta(t)} \leq \frac{1}{\sqrt{1-\eta}}.$$

Thus,

$$\lim_{t\to\infty} \frac{f(t)}{\sqrt{2}\tau\sqrt{\log t}} = \lim_{t\to\infty} \frac{f(t)}{(2\sqrt{2}/\sigma)\sqrt{\log t}} = 1$$

so with probability 1,

$$\lim_{t\to\infty} \frac{\ell_t}{(2\sqrt{2}/\sigma)\sqrt{\log t}} = \lim_{t\to\infty} \frac{\ell_t}{f(t)} \cdot \frac{f(t)}{(2\sqrt{2}/\sigma)\sqrt{\log t}} = 1.$$

□

To prove Theorem 8, we will need two lemmas. The first is general, and will be used several times in the sequel, while the second deals exclusively with the Gaussian case.

Denote by E_t the event that $a_\tau = +1$ for all $\tau \geq t$; that is, that there are no more mistakes after time t. The next lemma provides a uniform bound for the probability of E_t, conditioned on the public belief. It implies, in particular, that the probability of E_1 is positive, which we will use in the proof of Theorem 8.

Lemma 13. *Suppose G_- and G_+ are continuous, and G_- has a convex and differentiable left tail. Then for every $L \in \mathbb{R}$, there is some $m_L > 0$ such that for any t, $x \geq L$ implies $\mathbb{P}_+(E_t \mid \ell_t = x) \geq m_L$.*

Proof. Recall the definition of the public belief $\mu_t = \mathbb{P}(\theta = +1|a_1, \ldots, a_{t-1})$. The process $(\mu_1, \mu_2, \ldots)$ is a bounded martingale, and therefore, by a standard argument on bounded martingales, if we condition on $\mu_t = q$, then the probability that $\mu_\tau \leq 1/2$ for some $\tau > t$ is at most $2(1-q)$.[11] This event is precisely the complement of E_t, and therefore we have that $\mathbb{P}(E_t \mid \mu_t = q)$ is at least $2q - 1$. Hence, conditioning on $\theta = +1$, we have that $\mathbb{P}_+(E_t \mid \mu_t = 1-q) \geq (2q-1)/q$, which is positive for all $q > 1/2$.

Since $\mu_t = q$ is equivalent to $\ell_t = \log q/(1-q)$, what we have shown implies that there is an $\varepsilon > 0$ such that for all $x \geq 1$ (here the choice of 1 is arbitrary and can be

[11] Intuitively, if I assign high belief now to the event $\theta = +1$, then the probability that I assign this event low belief in the future must be small.

replaced with any positive number)

$$\mathbb{P}_+(E_t \mid \ell_t = x) > \varepsilon.$$

Now, for any $L < 1$, the compactness of the interval $[L, 1]$, together with the continuity of G_- and G_+, implies that there is an n_L such that if $\ell_t \geq L$, and if agents t through $t + n_L - 1$ take action +1, then $\ell_{t+n_L} > 1$. Further, since the probability of agents t through $t + n_L - 1$ all taking action +1 conditional on $\ell_t = x$ is continuous in x, there is a $p_L > 0$ such that

$$\mathbb{P}_+(E_t \mid \ell_t = x) \geq p_L \cdot \varepsilon$$

since with probability at least p_L there are no mistakes up to time $t + n_L$, and thence there are no mistakes with probability at least ε.

□

Lemma 14. *Assume private signals are Gaussian. For every $\varepsilon > 0$ there exists some $k > 0$ such that for all t,*

$$\mathbb{P}_+(a_t = -1 \mid a_\tau = +1 \textit{ for all } \tau < t) > \frac{k}{t^{1+\varepsilon}}.$$

Proof. By the definitions of ℓ_t^* and G_+,

$$\begin{aligned}\mathbb{P}_+(a_t = -1 \mid a_\tau = +1 \text{ for all } \tau < t) &= \mathbb{P}_+(a_t = -1 \mid \ell_t = \ell_t^*) \\ &= G_+(-\ell_t^*).\end{aligned}$$

Now, by Theorem 7, for every $\beta > 0$, $\ell_t^* < (1+\beta)\frac{2\sqrt{2}}{\sigma}\sqrt{\log t}$ for all sufficiently large t. Further, it follows from a routine application of L'Hopital's rule (or from the standard asymptotic expansion for the CDF of a normal distribution) that for all sufficiently large x,

$$G_+(-x) > \frac{\mathrm{e}^{-(\sigma^2/8)x^2}}{x}.$$

Let $\varepsilon > 0$, and take $\beta < \sqrt{1+\varepsilon} - 1$. Then by monotonicity of $G_+(-x)$ and a straightforward calculation,

$$\begin{aligned}G_+(-\ell_t^*) &> G_+(-(1+\beta)\frac{2\sqrt{2}}{\sigma}\sqrt{\log t}) \\ &> \left[\frac{1}{(1+\beta)\frac{2\sqrt{2}}{\sigma}}\right] \cdot \frac{t^{(1+\varepsilon)-(1+\beta)^2}}{\sqrt{\log t}} \cdot \frac{1}{t^{1+\varepsilon}} \\ &> \frac{1}{t^{1+\varepsilon}}\end{aligned}$$

for all sufficiently large t. From this, the claim follows immediately.

□

Proof of Theorem 8. Denote by C_t be the event that $a_\tau = +1$ for all $\tau < t$, and note that the event $T_1 = t$ is simply the intersection of C_t with the event that $a_t = -1$.

Let $\varepsilon > 0$. By Lemma 14 there is some $k' > 0$ such that for all t,

$$\mathbb{P}_+(a_t = -1 \mid C_t) > \frac{k'}{t^{1+\varepsilon}}.$$

Now, put $\gamma = \mathbb{P}_+(a_\tau = +1 \text{ for all } \tau \geq 1)$, the probability that all agents take the correct action. By Lemma 13, $\gamma > 0$, so this provides a lower bound on the probability of the first $t-1$ agents taking the correct action. Formally,

$$\mathbb{P}_+(C_t) \geq \mathbb{P}_+(a_\tau = +1 \text{ for all } \tau \geq 1) = \gamma.$$

Thus,

$$\begin{aligned}\mathbb{P}_+(T_1 = t) &= \mathbb{P}_+(a_t = -1, C_t)\\ &= \mathbb{P}_+(a_t = -1 \mid C_t) \cdot \mathbb{P}_+(C_t)\\ &\geq \frac{\gamma k'}{t^{1+\varepsilon}}\end{aligned}$$

for all t.

□

2.8 Upsets and runs

We recall a few definitions from Section 2.3. We say that there is an *upset* at time t if $a_{t-1} \neq a_t$. We denote by Ξ the random variable which assigns to each outcome the total number of upsets, and by Ξ_t the total number of upsets at times up to and including t. We say that there is a *run* of length m from t if $a_t = a_{t+1} = \cdots = a_{t+m-1}$. Note that this definition does not preclude a run from being part of a longer run; we will refer to a run of finite length which is not strictly contained in any other run as *maximal*. We say that a run from t is *good* if $a_t = +1$ and *bad* otherwise.

Notice that the number of maximal runs is exactly equal to the number of upsets. We use this observation now to show that the probability of having many maximal runs is very small, so that most of the probability is concentrated in the outcomes with few maximal runs.

Proof of Proposition 8. Denote by Υ the random variable which assigns to each outcome the number of finite maximal good runs it contains; note that with probability 1, Υ is finite.

By Lemma 13, there is a $\beta > 0$ such that for any $x \geq 0$, if $\ell_t = x$, then the probability that all agents from t on take the correct action is at least β. Formally,[12]

$$\mathbb{P}_+(a_\tau = +1 \text{ for all } \tau \geq t \mid \ell_t = x) \geq \beta.$$

Thus, whenever $a_{t-1} = -1$ and $a_t = +1$ (or $t = 1$), the probability that there is exactly one more maximal good run is at most $1 - \beta$. It follows that for $n \geq 0$,

$$\mathbb{P}_+(\Upsilon = n + 1) \leq (1 - \beta)\mathbb{P}_+(\Upsilon = n)$$

[12]We remind the reader that $\mathbb{P}_+(\cdot)$ is shorthand for $\mathbb{P}(\cdot \mid \theta = +1)$.

and thus, for any $n \geq 0$,

$$\mathbb{P}_+(\Upsilon = n) \leq (1-\beta)^n \mathbb{P}_+(\Upsilon = 0)$$

and so

$$\mathbb{P}_+(\Upsilon \geq n) \leq \frac{\mathbb{P}_+(\Upsilon = 0)}{\beta} \cdot (1-\beta)^n.$$

Finally, since $\Upsilon = \lfloor \Xi/2 \rfloor$, we have for any n:

$$\mathbb{P}_+(\Xi \geq n) \leq \mathbb{P}_+(\Upsilon \geq \lfloor n/2 \rfloor) \leq c \cdot \gamma^n$$

where $c = \mathbb{P}_+(\Upsilon = 0)/\beta$ and $\gamma = (1-\beta)^{\frac{1}{3}}$.

□

Whenever asymptotic learning occurs (that is, whenever the probability that all agents take the correct action from some point on is equal to 1), the total number of upsets is almost surely finite. In particular, the probability that Ξ_t is logarithmic in t tends to zero as t tends to infinity. Using Proposition 8, we can show that in fact this probability tends to 0 quickly:

Corollary 2. *Let c, γ be as in Proposition 8. Then*

$$\mathbb{P}(\Xi_t \geq -\frac{2.1}{\log \gamma} \log t) \leq c \cdot \frac{1}{t^{2.1}}.$$

Proof.

$$\begin{aligned}
\mathbb{P}(\Xi_t \geq -\frac{2.1}{\log \gamma} \log t) &\leq \mathbb{P}(\Xi \geq -\frac{2.1}{\log \gamma} \log t) \\
&\leq c \cdot \gamma^{-\frac{2.1}{\log \gamma} \log t} \\
&= c \cdot \frac{1}{t^{2.1}}.
\end{aligned}$$

□

In fact, it is equally easy to show the same statement for exponents larger than 2.1, but this will suffice for our purposes.

One important consequence of Corollary 2 is that with high probability, there is at least one maximal run before time t which is long relative to t. Thus, much of the dynamics is controlled by what happens during long runs.

We previously analyzed only long runs that start at time 1, when the public log-likelihood ratio is equal to 0. If a long run starts at some public belief $\ell_t \neq 0$ then its evolution is different from the former case. However, if the run is long enough then the analysis above can still be applied. The following lemma states that if a run starts at some $\ell_t > 0$ then we can bound the future public belief from below using ℓ^*.

Lemma 15. *Suppose that G_- has a convex and differentiable left tail. Then there exists a $z > 0$ such that, if there is a good run of length s from t, then $\ell_{t+s} \geq \ell^*_{s-z}$.*

Proof. Let $u_+(x) = x + D_+(x)$. Then by (2.1), whenever agent t takes action +1, $\ell_{t+1} = u_+(\ell_t)$.

Since G_- is eventually convex and differentiable, $u_+(x)$ is monotone increasing for sufficiently large x, by Lemma 12. Take

$$z = \min\{t \in \mathbb{N} \colon u_+(x) \text{ is monotone on } (\ell^*_t - 1, \infty)\}.$$

Now, let $\mu = \inf_{x \in [0,\ell^*_z]} D_+(x)$. By continuity of $D_+(x)$ and compactness of $[0, \ell^*_z]$, $\mu > 0$, since $D_+(x) > 0$ for all x. Put $N = \lceil \frac{\ell^*_z}{\mu} \rceil$. Then for all $x \in [0, \ell^*_z]$, $u_+^N(x) \geq \mu \cdot N \geq \ell^*_z$. Further, since $u_+(x) > x$ for all x, it follows that whenever there is a run of length N from t, $\ell_{t+N} > \ell^*_z$.

This implies that if there is a good run from t of length $s \geq N$, then $\ell_{t+s} \geq \ell^*_{s-z}$.

□

2.9 Distributions with polynomial tails

In this appendix we prove Theorem 9, showing that for private log-likelihood distributions with polynomial tails, the expected time to learn is finite.

As in the setting of Theorem 9, assume that the conditional distributions of the private log-likelihood ratio satisfy

$$G_+(x) = 1 - \frac{c}{x^k} \text{ for all } x > x_0 \tag{2.9}$$

$$G_-(x) = \frac{c}{(-x)^k} \text{ for all } x < -x_0 \tag{2.10}$$

for some $x_0 > 0$.

We remind the reader that we denote by ℓ^*_t the log-likelihood ratio of the public belief that results when the first $t - 1$ agents take action +1. It follows from Theorem 6 that in this setting, ℓ_t^* behaves asymptotically as $t^{1/(k+1)}$. Notice also that, by the symmetry of the model, the log-likelihood ratio of the public belief that results when the first $t-1$ agents take action -1 is $-\ell_t^*$.

We begin with the simple observation that a strong enough bound on the probability of mistake is sufficient to show that the expected time to learn is finite. Formally, we have the following lemma. We remind the reader that $\mathbb{P}_+(\cdot)$ is shorthand for $\mathbb{P}(\cdot \mid \theta = +1)$.

Lemma 16. *Suppose there exist $k, \varepsilon > 0$ such that for all $t \geq 1$, $\mathbb{P}_+(a_t = -1) < k \cdot \frac{1}{t^{2+\varepsilon}}$. Then $\mathbb{E}_+(T_L)$ is finite.*

Proof. Since $T_L = t$ only if $a_{t-1} = -1$, $\mathbb{P}_+(T_L = t) \leq \mathbb{P}_+(a_{t-1} = -1)$. Thus

$$\begin{aligned}
\mathbb{E}_+(T_L) &= \sum_{t=1}^{\infty} t \cdot \mathbb{P}_+(T_L = t) \\
&\leq \mathbb{P}_+(T_L = 1) + \sum_{t=2}^{\infty} t \cdot \mathbb{P}_+(a_{t-1} = -1) \\
&\leq 1 + k \sum_{i=2}^{\infty} \frac{t}{(t-1)^{2+\varepsilon}} \\
&< \infty.
\end{aligned}$$

□

Accordingly, this section will be primarily devoted to studying the rate of decay of the probability of mistake, $\mathbb{P}_+(a_t = -1)$. In order to bound this probability, we will need to make use of the following lemmas, which give some control over how the public belief is updated following an upset.

Lemma 17. *For G_+ and G_- as in* (2.9) *and* (2.10)*, $|\ell_{t+1}| \leq |\ell_t|$ whenever $|\ell_t|$ is sufficiently large and $a_t \neq a_{t+1}$.*

Proof. Assume without loss of generality that $a_t = +1$ and $a_{t+1} = -1$, so that

$$\ell_{t+1} = \ell_t + D_-(\ell_t).$$

Thus, to prove the claim we compute a bound for D_-. To do so we first obtain a bound for the left tail of G_+. By assumption, for $x > x_0$ (with x_0 as in (2.9) and (2.10)),

$$g_-(-x) = G'_-(-x) = \frac{ck}{x^{k+1}}$$

and so by (2.4),

$$g_+(-x) = \mathrm{e}^{-x} g_-(-x) = ck\frac{\mathrm{e}^{-x}}{x^{k+1}}.$$

Hence,

$$G_+(-x) = \int_{-\infty}^{-x} g_+(\zeta)\,\mathrm{d}\zeta = \int_{-\infty}^{-x} ck\frac{\mathrm{e}^{\zeta}}{(-\zeta)^{k+1}}\,\mathrm{d}\zeta = ck\int_{x}^{\infty} \zeta^{-k-1}\mathrm{e}^{-\zeta}\,\mathrm{d}\zeta.$$

For ζ sufficiently large, ζ^{-k-1} is at least, say, $\mathrm{e}^{-.1\zeta}$. Thus, for x sufficiently large,

$$G_+(-x) \geq ck\int_{x}^{\infty} \mathrm{e}^{-1.1\zeta}\,\mathrm{d}\zeta = \frac{ck}{1.1}\mathrm{e}^{-1.1x}.$$

It follows that for x sufficiently large,

$$D_-(x) = \log\frac{G_+(-x)}{G_-(-x)} \geq \log\frac{ck}{1.1} - 1.1x + k\log x \geq -1.2x.$$

Thus, for ℓ_t sufficiently large,

$$\ell_{t+1} = \ell_t + D_-(\ell_t) = \ell_t + \log\frac{G_+(-\ell_t)}{G_-(-\ell_t)} \geq \ell_t + 1.2(-\ell_t) = -.2\ell_t$$

so in particular, $|\ell_{t+1}| < |\ell_t|$.

□

We will make use of the following lemma, which bounds the range of possible values that ℓ_t can take.

Lemma 18. *For G_+ and G_- as in* (2.9) *and* (2.10)*, there exists an $M > 0$ such that for all $t \geq 0$, $|\ell_s| \leq M \cdot \ell_t^*$ for all $s \leq t$.*

Proof. For each $\tau \geq 0$, define

$$M_\tau = \max\frac{|\ell_\tau|}{\ell_\tau^*}$$

where the maximum is taken over all outcomes. Note that there are at most 2^τ possible values for this expression, so M_τ is well-defined and finite. Put

$$M = \sup_{\tau\geq 0} M_\tau.$$

To establish the claim, we must show that M is finite. To do this, it suffices to show that for τ sufficiently large, $M_{\tau+1} \leq M_\tau$.

Now, let $u_+(x) = x + D_+(x)$ and $u_-(x) = x + D_-(x)$. Then as shown in the section about the model, whenever agent τ takes action +1, $\ell_{\tau+1} = u_+(\ell_\tau)$, and whenever agent τ takes action −1, $\ell_{\tau+1} = u_-(\ell_\tau)$.

By Lemma 12, u_+ and u_- are eventually monotonic. Thus, there exists $x_0 > 0$ such that u_+ is monotone increasing on (x_0, ∞) and u_- is monotone decreasing on $(-\infty, -x_0)$.

For τ sufficiently large, $\ell^*_\tau > x_0$. Further, it follows from Lemma 17 that for τ sufficiently large, $|\ell_{\tau+1}| < |\ell_\tau|$ whenever $a_\tau \neq a_{\tau+1}$ and $|\ell_\tau| > |\ell^*_\tau|$. Let (a_τ) be any sequence of actions with $\frac{|\ell_{\tau+1}|}{\ell^*_{\tau+1}} = M_{\tau+1}$. If $a_\tau \neq a_{\tau+1}$

$$M_{\tau+1} = \frac{|\ell_{\tau+1}|}{\ell^*_{\tau+1}} \leq \frac{|\ell_\tau|}{\ell^*_\tau} \leq M_\tau.$$

If $a_\tau = a_{\tau+1}$, then either $M_{\tau+1} = 1$, in which case $M_{\tau+1} \leq M_\tau$, or $M_{\tau+1} > 1$. If $M_{\tau+1} > 1$, then since $|D_+|$ and $|D_-|$ are decreasing on (x_0, ∞) and $(-\infty, -x_0)$, respectively, $|\ell_{\tau+1} - \ell_\tau|/|\ell_\tau| \leq |\ell^*_{\tau+1} - \ell^*_\tau|/|\ell^*_\tau|$. So

$$M_{\tau+1} = \frac{|\ell_{\tau+1}|}{\ell^*_{\tau+1}} = \frac{|\ell_\tau| + |\ell_{\tau+1} - \ell_\tau|}{\ell^*_\tau + |\ell^*_{\tau+1} - \ell^*_\tau|}$$

where the second equality follows from the fact that ℓ_τ and $\ell_{\tau+1}$ have the same sign. Finally,

$$M_{\tau+1} = \frac{|\ell_\tau|}{\ell^*_\tau} \cdot \frac{1 + |\ell_{\tau+1} - \ell_\tau|/|\ell_\tau|}{1 + |\ell^*_{\tau+1} - \ell^*_\tau|/\ell^*_\tau} \leq \frac{|\ell_\tau|}{\ell^*_\tau} \leq M_\tau.$$

Thus, for all sufficiently large τ, $M_{\tau+1} \leq M_\tau$.

□

Proposition 9. *There exists* $\kappa > 0$ *such that* $\mathbb{P}_+(a_t = -1) < \kappa t^{-2.1}$ *for all* $t > 0$.

Proof. Let $\beta = -2.1/\log\gamma$, where γ is as in Proposition 8. To carry out our analysis, we will divide the event that $a_t = -1$ into three disjoint events and bound each of them separately:

$$A = (a_t = -1) \text{ and } (\Xi_t > \beta \log t)$$
$$B_1 = (a_t = -1) \text{ and } (\Xi_t \leq \beta \log t) \text{ and } (|\{s\colon s < t, a_s = +1\}| \geq \frac{1}{2}t)$$
$$B_2 = (a_t = -1) \text{ and } (\Xi_t \leq \beta \log t) \text{ and } (|\{s\colon s < t, a_s = +1\}| < \frac{1}{2}t).$$

First, by Corollary 2 we have a bound for $\mathbb{P}_+(A)$

$$\mathbb{P}_+(A) \leq c \cdot \frac{1}{t^{2.1}}.$$

Next, we bound $\mathbb{P}_+(B_1)$. This is the event that the number of upsets so far is small and the majority of agents so far have taken the correct action.

Since there are at most $\beta \log t$ upsets, there are at most $\frac{1}{2}\beta \log t$ maximal good runs. Since, furthermore, there are at least $\frac{1}{2}t$ agents who take action +1, there is at least one maximal good run of length at least $t/(\beta \log t)$.

Thus, $\mathbb{P}_+(B_1)$ is bounded from above by the probability that there are some $s_1 < s_2 < t$ such that there is a good run of length $s_2 - s_1 \geq t/(\beta \log t)$ from s_1 and $a_{s_2} = -1$.

For fixed s_1, s_2, denote by E_{s_1,s_2} the event that there is a good run of length $s_2 - s_1$ from s_1. Denote by Γ_{s_1,s_2} the event $(E_{s_1,s_2}, a_{s_2} = -1)$. Then

$$\begin{aligned}\mathbb{P}_+(\Gamma_{s_1,s_2}) &= \mathbb{P}_+(a_{s_2} = -1 | E_{s_1,s_2}) \cdot \mathbb{P}_+(E_{s_1,s_2}) \\ &\leq \mathbb{P}_+(a_{s_2} = -1 | E_{s_1,s_2}).\end{aligned}$$

By Lemma 15, there exists a $z > 0$ such that E_{s_1,s_2} implies that $\ell_{s_2} \geq \ell^*_{s_2-s_1-z}$. Therefore,

$$\mathbb{P}_+(\Gamma_{s_1,s_2}) \leq G_+(-\ell^*_{s_2-s_1-z}).$$

Since for t sufficiently large $\ell^*_t > t^{\frac{1}{k+2}}$ and since $G_+(-x) \leq \mathrm{e}^{-x}$ by (2.4),

$$\mathbb{P}_+(\Gamma_{s_1,s_2}) \leq \mathrm{e}^{-\alpha(s_2-s_1-z)^{\frac{1}{k+2}}} \leq \mathrm{e}^{-\alpha(t/(\beta \log t)-z)^{\frac{1}{k+2}}}.$$

To simplify, we further bound this last expression to arrive at, for some $c > 0$,

$$\mathbb{P}_+(\Gamma_{s_1,s_2}) \leq c\mathrm{e}^{-t^{\frac{1}{k+3}}}$$

for all t. Since B_1 is covered by fewer than t^2 events of the form Γ_{s_1,s_2} (as s_1 and s_2 are less than t), it follows that

$$\mathbb{P}_+(B_1) < ct^2\mathrm{e}^{-t^{\frac{1}{k+3}}} < \frac{1}{t^{2.1}}$$

for all t large enough.

Finally we bound $\mathbb{P}_+(B_2)$. This is the event that the number of upsets so far is small and the majority of agents so far have taken the wrong action. As in B_1, there is a maximal bad run of length at least $t/(\beta \log(t))$.

Denote by R the event that there is at least one bad run of length $t/(\beta\log(t))$ before time t and by R_s the event that agents s through $s+t/(\beta\log t)-1$ take action -1. Since B_2 is contained in R, and since R is contained in the union $\cup_{s=1}^{t} R_s$, we have that

$$\mathbb{P}_+(B_2) \le \mathbb{P}_+(R) \le \sum_{s=1}^{t} \mathbb{P}_+(R_s).$$

Taking the maximum of all the addends in the right hand side, we can further bound the probability of B_2:

$$\mathbb{P}_+(B_2) \le t \cdot \max_{1\le s\le t} \mathbb{P}_+(R_s).$$

Conditioned on ℓ_s, the probability of R_s is

$$\mathbb{P}_+(R_s \,|\, \ell_s) = \prod_{r=s}^{s+t/(\beta\log t)-1} G_+(-\ell_r).$$

By Lemma 18, there exists $M > 0$ such that $|\ell_r| \le M\ell_t^*$, for all $r \le t$. Therefore, since G_+ is monotone,

$$\mathbb{P}_+(R_s) \le G_+(M\ell_t^*)^{t/(\beta\log t)}.$$

It follows that

$$\mathbb{P}_+(B_2) \le t \cdot G_+(M\ell_t^*)^{t/(\beta\log t)}.$$

Since $G_+(x) = 1 - c\cdot x^{-k}$ for x large enough, and since ℓ_t^* is asymptotically at most $t^{1/(k+0.5)}$, we have that

$$\log G_+(M\ell_t^*) \le -cM^{-k}\cdot t^{-k/(k+0.5)}.$$

Thus

$$\mathbb{P}_+(B_2) \le t\cdot\exp\left(-cM^{-k}\cdot t^{1/(2k+1)}/(\beta\log t)\right) \le t^{-2.1},$$

for all t large enough. This concludes the proof, because $\mathbb{P}_+(a_t = -1) = \mathbb{P}_+(A) + \mathbb{P}_+(B_1) + \mathbb{P}_+(B_2) \le \kappa\frac{1}{t^{2.1}}$ for some constant κ.

□

Given this bound on the probability of mistakes, the proof of the main theorem of this section follows easily from Lemma 16.

Proof of Theorem 9. By Proposition 9, there exists $\kappa > 0$ such that $\mathbb{P}(a_t = -1 \mid \theta = +1) < \kappa \frac{1}{t^{2.1}}$ for all $t \geq 1$. Hence, by Lemma 16 $\mathbb{E}(T_L \mid \theta = +1) < \infty$. By a symmetric argument the same holds conditioned on $\theta = -1$. Thus, the expected time to learn is finite. □

Chapter 3

EQUITABLE VOTING RULES

3.1 Introduction

Literally translated to "power of the people," democracy is commonly associated with two fundamental tenets: equity among individuals and responsiveness to their choices. May's celebrated theorem provides foundation for voting systems satisfying these two restrictions May (1952). Focusing on two-candidate elections, May illustrated that majority rule is unique among voting rules that treat candidates identically and guarantee symmetry and responsiveness.

Extensions of May's original results are bountiful.[1] However, what we view as a procedural equity restriction in his original treatment—often termed anonymity or symmetry—has remained largely unquestioned.[2] This restriction requires that no two individuals can affect the collective outcome by swapping their votes. Motivated by various real-world voting systems, this paper focuses on a particular weakening of this restriction. While still capturing the idea that no voter carries a special role, our equity notion allows for a large spectrum of voting rules, some of which are used in practice, and some of which we introduce. We analyze winning coalitions of such equitable voting rules and show that they can comprise a vanishing fraction of the population, but not less than the square root of its size. Methodologically, we demonstrate how techniques from group theory can be useful for the analysis of fundamental questions in social choice.

To illustrate our motivation, consider the stylized example of a *representative democracy* rule: *m counties* each have *k* residents. Each county selects, using majority rule, one of two representatives. Then, again using majority rule, the *m* representatives select one of two policies (see Figure 3.1 for the case $m = k = 3$).

This rule does not satisfy May's original symmetry restriction: individuals could swap their votes and change the outcome. In Figure 3.1, for example, suppose voters $\{1, 2, 3, 4, 5\}$ vote for representatives supporting policy A, while voters $\{6, 7, 8, 9\}$

[1] See, e.g., Cantillon and Rangel (2002), Fey (2004), Goodin and List (2006), and references therein.

[2] An exception is Packel (1980), who relaxes the symmetry restriction and adds two additional restrictions to generate a different characterization of majority rule than May's.

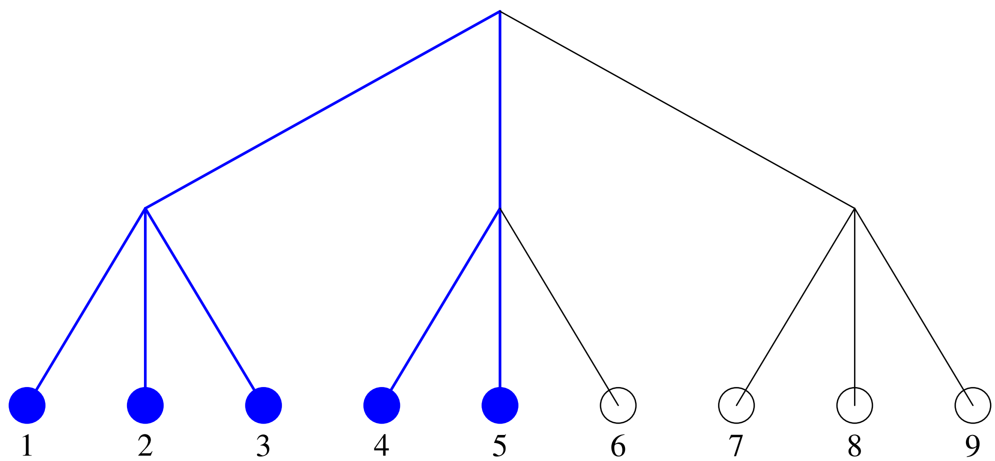

Figure 3.1: A representative democracy voting rule. Voters are grouped into three counties: $\{1,2,3\}$, $\{4,5,6\}$ and $\{7,8,9\}$. Each county elects a representative by majority rule, and the election is decided by majority rule of the representatives.

vote for representatives supporting policy B. With the original votes, policy A would win; but swapping voters 5 and 9 would cause policy B to win.

Even though representative democracy rules do not satisfy May's symmetry assumption, there certainly is an intuitive sense in which their fundamental characteristics "appear" equitable. Indeed, variations of these rules were chosen in good faith by many designers of modern democracies. Such rules are currently in use in France, India, the United Kingdom, and the United States, among others.

Is there a formal sense in which a representative democracy rule is more equitable than a dictatorship? More generally, what makes a voting rule equitable? We suggest the following definition. In an *equitable voting rule*, for any two voters v and w, there is some permutation of the full set of voters such that: (i) the permutation sends v to w, and (ii) applying this permutation to any voting profile leaves the election result unchanged.

In §3.2, we formalize the notion of "roles" in a voting body. For instance, in university committees there are often two distinct roles: a chair and a standard committee member; likewise, in juries, there is often a foreperson, who carries a special role, and several jury members, who all have the same role. We show that our notion of equity is tantamount to *all* agents in the electorate having the same role.

Under our equity definition, representative democracy rules are indeed equitable, but dictatorships are not. For instance, in the case depicted in Figure 3.1, voters 1

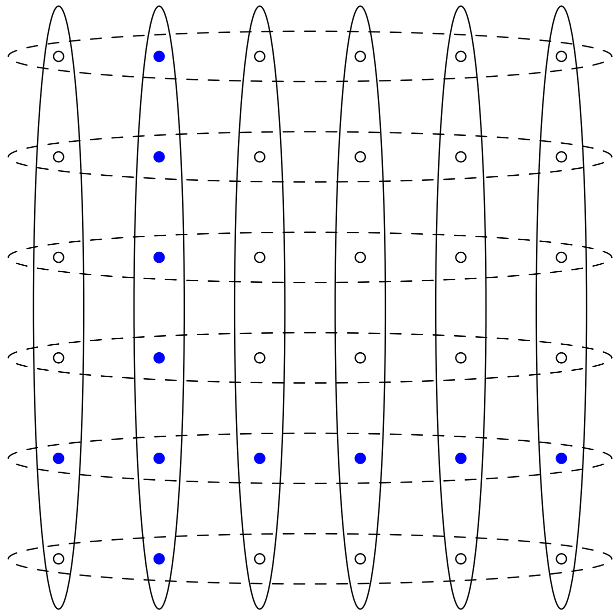

Figure 3.2: Cross-committee consensus voting rule. The union of a row and a column is a winning coalition.

and 2 play the same role, since the permutation that swaps them leaves any election result unchanged. But 1 and 4 also play the same role: the permutation that swaps the first *county* with the second *county* also leaves outcomes unchanged.

There is a large variety of equitable rules that are not representative democracy rules. An example is what we call *Cross Committee Consensus (CCC)* rules. In these, each voter is assigned to two committees: a “row committee” and a “column committee” (see Figure 3.2). If any row committee and any column committee both exhibit consensus, then their choice is adopted. Otherwise, majority rule is followed. For instance, suppose a university is divided into equally-sized departments, and each faculty member sits on one university-wide committee. CCC corresponds to a policy being accepted if there is a strong unanimous lobby from a department and from a university-wide committee, with majority rule governing decisions otherwise. This rule is equitable since each voter is a member of precisely one committee of each type, and all row (column) committees are interchangeable.

We provide a number of further examples of equitable voting rules, showing the richness of this class and its versatility in allowing different segments of society—counties, university departments, etc.—to express their preferences.

In order to characterize more generally the class of equitable voting rules, we focus on their *winning coalitions*, the sets of voters that decide the election when in agreement Reiker (1962). In majority rule, all winning coalitions include at least half of the population. We analyze how small winning coalitions can be in equitable voting rules.

Under representative democracy, winning coalitions have to comprise at least a quarter of the population.[3] Much smaller winning coalitions are possible in what we call *generalized representative democracy (GRD)* voting rules, where voters are hierarchically divided into sets that are, in turn, divided into subsets, and so on. For each set, the outcome is given by majority rule over the decisions of the subsets.[4] We show that equitable GRD rules for n voters can have winning coalitions as small as $n^{\log_3 2}$, or about $n^{0.63}$, which is a vanishingly small fraction of the population.[5]

When the number of voters n is a perfect square, and when committee sizes are taken to be $\sqrt{n}$, the CCC rule has a winning coalition of size $2\sqrt{n} - 1$. This is significantly smaller than $n^{\log_3 2}$, for n large enough.

Our main result is that, for any n, there always exist simple equitable voting rules that have winning coalitions of size $\approx 2\sqrt{n}$. Conversely, we show that no equitable voting rule can have winning coalitions of size less than $\sqrt{n}$. Methodologically, the proof utilizes techniques from group theory and suggests the potential usefulness of such tools for the analysis of collective choice.

While $\sqrt{n}$ accounts for a vanishing fraction of the voter population, we stress that it can be viewed as "large" in many contexts. While in a department of 100 faculty, 10 members would need to coordinate to sway a decision one way or the other, in a presidential election with, say, 140 million voters, coordination between nearly $12,000$ voters would be necessary to impact outcomes.[6]

[3] A winning coalition under representative democracy must include support from half the counties, which translates to half of the population in those counties, or one quarter of the entire population.

[4] These rules have been studied under the name *recursive majority* in the probability literature (see, e.g., Mossel and O'Donnell, 1998).

[5] For an example of a non-equitable GRD with a small winning coalition, consider voters $\{1, \ldots, 1000\}$ and assume three counties divide the population into three sets of voters: $\{\{1\}, \{2\}, \{3, \ldots, 1000\}$. Then $\{1, 2\}$ is a winning coalition. The value $\log_3 2 \approx 0.63$ is the Hausdorff dimension of the Cantor set. As it turns out, there is a connection between equitable GRD's that achieve the smallest winning coalitions and the Cantor set.

[6] Interestingly, rules that give decisive power to minorities of size $\sqrt{n}$ appear in other contexts of collective choice and have been proposed for apportioning representation in the United Nations Parliamentary Assembly, and for voting in the Council of the European Union, see Zyczkowski and Slomczynski (2014). These proposed rules relied on the Penrose Method Penrose (1946), which

For instance, even under majority rule, if each voter is equally likely to vote for either of two alternatives, the Central Limit Theorem suggests that a coalition of order $\sqrt{n}$ can control the vote with high probability.

Certainly, beyond equity, another important aspect of voting rules is their susceptibility to manipulation. For instance, with information on voters' preferences, representative democracy rules are sensitive to gerrymandering McGann et al. (2016). We view the question of manipulability as distinct from that of equity. It would be interesting to formulate a notion of non-manipulability, independent of equity, and to understand how these notions interact. The breadth of equitable voting rules allows for further consideration of various objectives, such as non-manipulability, when designing institutions.

In the next part of the paper, we explore a stronger notion of equity. We consider k-*equitable* voting rules in which every coalition of k voters plays the same role. These are increasingly stringent conditions that interpolate between our equity notion, when $k = 1$, and May's symmetry, when $k = n$. The analysis of k-equitable rules is delicate, due to group- and number-theoretical phenomena. There do exist, for arbitrarily large population sizes n, voting rules that are 2- and 3-equitable, and have winning coalitions as small as $\sqrt{n}$. However, for "most" sufficiently large values of n, and for any $k \geq 2$, the *only* k-equitable, neutral, and responsive voting rule is majority. Thus, while equity across individuals allows for a broad spectrum of voting rules, equity among arbitrary fixed-size coalitions usually places the restrictions May had suggested. While k-equity is arguably a strong restriction, it is still far weaker than May's original symmetry requirement. In that respect, our results here provide a strengthening of May's conclusions.

suggests the vote weight of any representative should be the square root of the size of the population she represents, when majority rule governs decisions. Penrose argued that this rule assures equal voting powers among individuals.

3.2 The model

Voting rules

Let V be a finite set of voters. We denote $V = \{1, \ldots, n\}$ so that n is the number of voters. Each voter has preferences over alternatives in the set $Y = \{-1, 1\}$. We identify the possible preferences over Y with elements of $X = \{-1, 0, 1\}$, where -1 represents a strict preference for -1 over 1, 1 represents a strict preference for 1 over -1, and 0 represents indifference between -1 and 1. We denote by $\Phi = X^V$ the set of voting profiles; that is, Φ is the set of all functions from the set of voters V to the set of possible preferences X. A *voting rule* is a function $f : \Phi \to X$.

An important example is the *majority* voting rule $\mathrm{m} \colon \Phi \to X$, which is given by

$$\mathrm{m}(\phi) = \begin{cases} 1 & \text{if } |\phi^{-1}(1)| > |\phi^{-1}(-1)| \\ -1 & \text{if } |\phi^{-1}(1)| < |\phi^{-1}(-1)| \\ 0 & \text{otherwise.} \end{cases}$$

A vote of 0 can be interpreted as abstention or indifference.

May's Theorem

We now define several properties of voting rules. Following May (1952), we say that a voting rule f is *neutral* if $f(-\phi) = -f(\phi)$. Neutrality implies that both alternatives -1 and 1 are treated symmetrically: if each individual flips her vote, the final outcome is also flipped.

Again following May (1952), we say that a voting rule f is *positively responsive* if increased support for one alternative makes it more likely to be selected. Formally, f is positively responsive if $f(\phi) = 1$ whenever there exists a voting profile ϕ' satisfying the following:

1. $f(\phi') = 0$ or 1.
2. $\phi(v) \geq \phi'(v)$ for all $v \in V$.
3. $\phi(v_0) > \phi'(v_0)$ for some $v_0 \in V$.

Thus, $f(\phi) \geq f(\phi')$ if $\phi \geq \phi'$ coordinate-wise, and if $f(\phi) = 0$ then any change of ϕ breaks the tie.

We now turn to the symmetry between voters. Several group-theoretic concepts will prove useful for the description and comparison of May's and our notions. Denote by S_n the set of permutations of the n voters. Any permutation of the voters can be associated with a permutation of the set of voting profiles Φ: given a permutation $\sigma \in S_n$, the associated permutation on the voting profiles maps ϕ to ϕ^σ, which is given by $\phi^\sigma(v) = \phi(\sigma^{-1}v)$. The *automorphism group* of the voting rule f is given by

$$\mathrm{Aut}_f = \{\sigma \in S_n \,|\, \forall \phi \in \Phi,\ f(\phi^\sigma) = f(\phi)\}.$$

That is, Aut_f is the set of permutations of the voters that leave election results unchanged, *for every* voting profile.

We can interpret a permutation σ as a scheme in which each voter v, instead of casting her own vote, gets to decide how some *other voter* $w = \sigma(v)$ will vote. A permutation σ is in Aut_f if applying this scheme never changes the outcome: when each $w = \sigma(v)$ votes as v would have, the result is the same as when each player v votes for herself.

The automorphism group Aut_f has natural implications for pivotality, or the Shapley-Shubik and Banzhaf indices of players in simple games, see Dubey and Shapley (1979). Consider a setting in which all voters choose their votes identically and independently at random. Given such a distribution, we can consider the probability η_v that a voter v is pivotal.[7] It is easy to see that if there is some $\sigma \in \mathrm{Aut}_f$ that maps v to w, then $\eta_v = \eta_w$, implying that v and w have the same Banzhaf index. In fact, when there exists $\sigma \in \mathrm{Aut}_f$ that maps v to w, any statistic associated with a voter that treats other voters identically—the probability the outcome coincides with voter v's vote, the probability that voter v and another voter are pivotal, etc.—would be the same for voters v and w. Hence, in an equitable rule, all the voters' Banzhaf indices will be equal.[8]

May (1952)'s notion of equity, often termed *symmetry* or *anonymity*, requires that swapping the votes of any two voters will not affect the collective outcome. It can be succinctly stated as $\mathrm{Aut}_f = S_n$.

May's Theorem. *Majority rule is the unique symmetric, neutral, and positively responsive voting rule.*

[7] A voter v is pivotal at a particular voting profile if a change in her vote can affect the outcome under f.

[8] In a recent follow-up paper to this paper, Bhatnagar (2020) shows that the converse does not hold: there are rules that are not equitable, but for which the same holds.

Perhaps surprisingly, the requirement of symmetry is stronger than what is needed for May's conclusions. As it turns out, a weaker requirement that Aut_f be restricted to only *even permutations* would suffice for his results, see Lemma 21 in the appendix.[9] In Theorem 14 below we show that, in fact, a far weaker requirement suffices.

Equitable Voting Rules

As we have already seen, the requirement that Aut_f includes all permutations, or all even permutations, precludes many examples of voting rules that "appear" equitable. What makes a voting rule appear equitable? Our view is that, in an equitable voting rule, ex-ante, all voters carry the same "role." We propose the following definition and discuss in the next section the sense in which it formalizes this view.

Definition 1. *A voting rule f is equitable if for every $v, w \in V$ there is a $\sigma \in \text{Aut}_f$ such that $\sigma(v) = w$.*

In words, a voting rule is equitable if, for any two voters v and w, there is some permutation of the population that relabels v as w such that, *regardless of voters' preferences*, the outcome is unchanged relative to the original voter labeling.

In group-theoretic terms, f is equitable if and only if the group Aut_f *acts transitively* on the voters.[10] Insights from group theory related to the characteristics of transitive groups are therefore at the heart of our main results. Appendix 3.7 contains a short primer on the basic group theoretical background that is needed for our analysis. [11]

Equity as Role Equivalence

May noted the strong link between anonymity and equality, stating that "This condition might well be termed anonymity... A more usual label is equality" (May, 1952, page 681). Are anonymity and equality inherently one and the same? In this section, we formalize this question in terms of *roles*. This allows for a natural distinction between May's symmetry or anonymity condition and our equity notion.

[9] A permutation σ is even if the number of pairs (v, w) such that $v < w$ and $\sigma(v) > \sigma(w)$ is even. Put another way, define a transposition to be a permutation that only switches two elements, leaving the rest unchanged. A permutation is even if it is the composition of an even number of transpositions.

[10] It turns out our equity restriction is effectively the definition of transitivity. The notion of transitive groups is not directly related to transitivity of relations often considered in Economics.

[11] Isbell (1960) studied a notion equivalent to our equity notion and considered its implications, combined with rule neutrality, in a setting in which preferences must be strict, so that tie breaking must also be equitable. When n is odd, majority is equitable and neutral. Isbell showed that for some even n such rules exist, while for other even n they do not.

In particular, we formalize our motivating idea that equity corresponds to all voters carrying the same role.

Even though we defined voting rules with respect to a given set of voters, the design of voting rules is often carried out without a particular group of people in mind; rather, collective institutions are often designed in the abstract. For example, hiring protocols in university departments might be specified prior to any specific search. One such protocol might be that a committee chair decides dictatorially whom to hire, unless indifferent, in which case the committee decides by majority rule. Under such an abstract rule, the dictatorial privilege is not assigned to a particular Prof. X, but to an abstract role called "chair." Later, when a committee is formed, the role of chair is assigned to some particular faculty member. The same applies, at least in aspiration, to countries' election rules, jury decision protocols, etc. Indeed, historical cases in which laws were written with particular individuals specified or implied are often not benevolent examples of institution building.

To capture this idea, we introduce abstract voting rules. Recall that we define (non-abstract) voting rules as functions from X^V to X in the context of a particular voter set V. An *abstract* voting rule is a map $f\colon X^R \to X$ for a set of *roles* R. Of course, mathematically, these objects are identical, and so we can speak of abstract voting rules as being anonymous, equitable, etc. In the above example of the committee, the set of roles would be $R = \{C, M_2, \dots, M_n\}$, where C stands for "chair" and M_i is member i.

The conceptual difference between voters and roles is that voters have preferences and vote, whereas roles do not. Therefore, for a vote to take place, voters need to be assigned to roles. Accordingly, given a group of voters V equal in size to R, we call a bijection $a\colon V \to R$ a *role assignment*. Given an abstract voting rule $f\colon X^R \to X$, a role assignment a defines a (non-abstract) voting rule $f_a\colon X^V \to X$ in the obvious way, via $f_a(\phi) = f(\phi \circ a^{-1})$. In our university committee example, if $V = \{\text{Alex}, \text{Bailey}, \dots\}$, an assignment a that satisfies $a(\text{Alex}) = C$ and $a(\text{Bailey}) = M_7$ assigns Alex the role of chair, and Bailey the role of member 7. Hence, the voting rule f_a is a dictatorship of Alex. A different assignment b with $b(\text{Bailey}) = C$ and $b(\text{Alex}) = M_7$ results in the voting rule f_b in which Bailey is the dictator. Note that any assignment c that also assigns $c(\text{Bailey}) = C$ results in the same voting rule as b, even if, say, $c(\text{Alex}) = M_8$: $f_b(\phi) = f_c(\phi)$ for any voting profile ϕ. In the context of an abstract voting rule f, we say that two assignments are equivalent if they lead to the same voting rule: a and b are equivalent under f if $f_a = f_b$.

The next proposition captures a sense in which symmetry is a form of equality:

Proposition 1A. *An abstract voting rule $f\colon X^R \to X$ is symmetric if and only if all assignments are equivalent under f.*

Thus, symmetry means that assignments do not matter, or that voters are completely indifferent between assignments: given any voting profile ϕ and given f, each voter would be indifferent if given a choice between assignments. Voters do not care which role they have; moreover, voters do not care which roles other voters have.

We now turn to the interpretation of our equity notion in terms of roles. In the university committee example, certainly "chair" is a distinguished role. Likewise, it is clear that if we view the representative democracy example of Figure 3.1 as an abstract voting rule, no role is distinguished. Of course, once we assign roles to voters with particular preferences, some voters may be disadvantaged, and prefer a different assignment ex-post. In that respect, in the abstract representative democracy rule, all roles are ex-ante identical.

We say that roles $r_1, r_2 \in R$ are equivalent under an abstract voting rule $f\colon X^R \to X$ if, for any voter v and assignment a such that $a(v) = r_1$, there is an assignment b with $b(v) = r_2$ such that $f_a = f_b$. That is, roles r_1 and r_2 are equivalent if, for any voter, it is impossible to determine whether they are assigned the role of r_1 or r_2 from the entire mapping from vote profiles to chosen alternatives. In the university committee example, M_i and M_j are equivalent, but C is not equivalent to any other role. If $a(v) = C$, then clearly this can be determined from f_a, but not if $a(v) \neq C$; in the latter case one can find an assignment b such that $b(v) \neq a(v)$, but $f_a = f_b$, since it is impossible to tell whether a voter has role M_i or M_j. In the representative democracy example of Figure 3.1, it is impossible to determine from f_a the role of any given voter v.

Given this definition of equivalent roles, the next proposition is a sharp characterization of equity.

Proposition 1B. *An abstract voting rule $f\colon X^R \to X$ is equitable if and only if all roles are equivalent under f.*

Therefore, while symmetry means that voters are indifferent between assignments, equity implies that voters are indifferent between roles. Indeed, Proposition 1B implies that, given an abstract voting rule f, and given any two roles r_1, r_2, if a voter had to choose between (i) any assignment in which she had role r_1, or (ii) any assignment in which she had role r_2, she would be indifferent. Both would allow her to select the same voting rule f_a. Likewise, if the voter had to choose between roles r_1 and r_2 knowing that an adversary would get to choose the rest of the assignment, she would be indifferent.

The following highlights the idea that equity captures indifference between roles.

Proposition 2A. *An abstract voting rule $f : X^R \to X$ is equitable iff there is a set of assignments A such that*

1. $f_a = f_b$ *for all* $a, b \in A$, *and*
2. *for each role* $r \in R$ *and voter* $v \in V$ *there is an* $a \in A$ *with* $a(v) = r$.

Thus, there is a menu of assignments that all induce the same voting rule, but allow v to choose any role.

Winning Coalitions

One way to describe a voting rule is through its winning coalitions: the sets of individuals whose consensual vote determines the alternative chosen. Formally, we say that a subset $S \subseteq V$ is a *winning coalition* with respect to the voting rule f if, for every voting profile ϕ and $x \in \{-1, 1\}$, $\phi(v) = x$ for all $v \in S$ implies $f(\phi) = x$.

Note that no two winning coalitions of f can be disjoint. Indeed, suppose that $M, M' \subseteq V$ are two disjoint winning coalitions. We can then have a profile under which members of M vote unanimously for -1 and members of M' vote for 1. In such cases, f would not be well defined. The following lemma illustrates a version of the converse.

Lemma 19. *Let $\mathcal{W}$ be a collection of subsets of V such that every pair of subsets in $\mathcal{W}$ has a nonempty intersection. Then there is a neutral, positively responsive voting rule for V for which every set in $\mathcal{W}$ is a winning coalition.*

Intuitively, the construction underlying Lemma 19 is as follows. First, for any vote profile in which a subset $W \in \mathcal{W}$ votes for 1 (or -1) in consensus, we specify the voting rule to also take the value of 1 (or -1). For any profile in which no $W \in \mathcal{W}$ votes in consensus, we define the voting rule to follow majority rule. By definition, the winning coalitions of this voting rule contain the sets in $\mathcal{W}$. As we show, it is also neutral and positively responsive.

This lemma will allow us to discuss neutral and positive-responsive equitable voting rules through the restrictions they impose on winning coalitions.

3.3 Winning coalitions for equitable voting rules

In this section, we provide bounds on the size of winning coalitions in general equitable voting rules. We then restrict attention to the special class of equitable voting rules that generalize representative democracy rules and characterize the size of winning coalitions for those.

Winning Coalitions of Order $\sqrt{n}$

We first show that for any population size, there always exist equitable voting rules that have winning coalitions that are of order $\sqrt{n}$.

Theorem 10. *For every n there exists a neutral, positively responsive equitable voting rule with winning coalitions of size* $2\lceil \sqrt{n} \rceil - 1$.

An important implication of this theorem is that, under an equitable voting rule, winning coalitions can account for a vanishing fraction of the population. Nevertheless, $\sqrt{n}$ is arguably a large number of voters in some contexts. For example, in a majority vote, suppose all voters vote for each of $\{-1, 1\}$ independently with probability one half. A manipulator who wants to guarantee an outcome with high probability would need to control an order of $\sqrt{n}$ of the votes. This is a consequence of the fact that the standard deviation of the number of voters who vote 1 is of order $\sqrt{n}$.

The cross committee consensus rule described in the introduction is an example of an equitable voting rule in which winning coalitions are $O(\sqrt{n})$. Nevertheless, there is an algebraic subtlety—the construction of that rule relies on n being an integer squared. Certainly, an analogous construction can be made for any n that can be described as $n = k \cdot m$ for some integers k and m by considering some committees to be of size k and others to be of size m. Such constructions, however, would not necessarily generate voting rules with winning coalitions of size close to $\sqrt{n}$. We prove Theorem 10 by constructing a simple, related rule that applies to every n, called the *longest-run* rule.[12]

Identify the set of voters with $\{0, 1, \ldots, n-1\}$, and place them along a cycle, as in Figure 3.3. Given a voting profile ϕ, a *run* is a contiguous block of voters

[12] We thank Elchanan Mossel for suggesting this improvement to a previous construction.

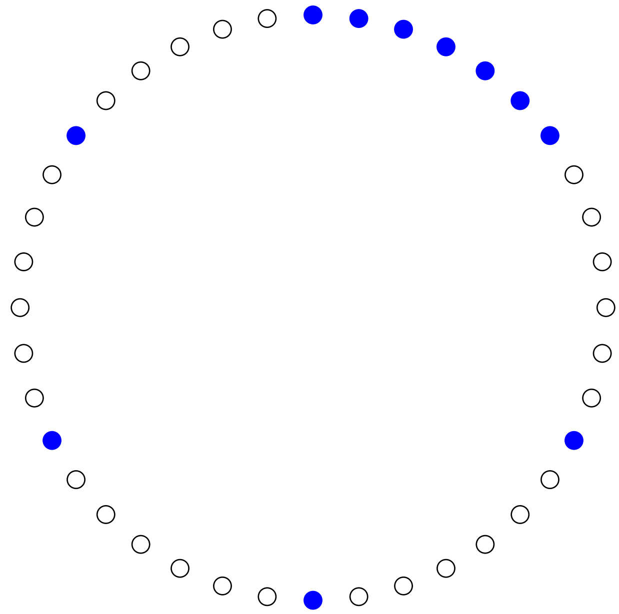

Figure 3.3: The longest-run voting rule. The set depicted in blue is a winning coalition of size $\approx 2\sqrt{n}$.

voting identically for either -1 or 1. Formally, a set of voters $W \subseteq V$ is a run if $\phi(w) = \phi(w') \in \{-1, 1\}$ for all $w, w' \in W$, and if $W = \{i, i+1, \ldots, i+k\}$ modulo n.

Given a voting profile ϕ, we say that W is the *longest-run* if it is a run that is strictly longer than all other runs. The longest-run voting rule ℓ is defined as follows: if there is a longest-run in ϕ, then $\ell(\phi)$ is the vote cast by the members of this run. Otherwise, $\ell(\phi) = \mathrm{m}(\phi)$, where m is the majority rule.

The longest-run rule is equitable, since the group of rotations maps any voter to any voter. Furthermore, it admits winning coalitions of size $\approx 2\sqrt{n}$: these include a run of length $\sqrt{n}$, together with $\sqrt{n}$ agents spaced $\sqrt{n}$ apart, thus preventing the creation of longer runs. See Figure 3.3.

We now offer a counterpart for Theorem 10 that provides a lower bound on the size of minimal coalitions in equitable voting rules.

Theorem 11. *Every winning coalition of an equitable voting rule has size at least* $\sqrt{n}$.

The proof of Theorem 11 relies on group-theoretic results described in Appendix 3.7. To gain some intuition for the bound, suppose, as in the longest-run voting rule above, that voters are located on a circle and that Aut_f includes all rotations. These are the permutations of the form $\sigma(i) = i + k \mod n$. We know that two winning coalitions cannot be disjoint. Take, then, any winning coalition S and denote by $S + k$ the winning coalition that is derived by adding k (again, modulo n) to the label of each member. It follows that S and $S + k$ must have a non-empty intersection, or that there are some $i, j \in S$ with $i - j = k$. Therefore, if we look at all the differences between two elements of S (i.e., expressions of the form $i - j$, where $i, j \in S$), they encompass all n rotations. In particular, the cardinality of these differences is n. On the other hand, the number of such differences is certainly bounded by the number of ordered pairs of members in S, which is $|S|^2$. It follows that $|S|^2 \geq n$, generating our bound.

Generalized Representative Democracy Rules

As already discussed, voting rules mimicking representative democracy are equitable, if not symmetric à la May (1952). We now consider a class of equitable voting rules that generalize representative democracy rules. These capture the flavor of various hierarchical voting structures that contain more than two layers. For example, voters may belong to counties, which comprise states, which constitute a country. As we show, these sorts of hierarchical decision rules are associated with far smaller winning coalitions than $n/2$, but still substantially larger than $\sqrt{n}$.

A voting rule $f\colon \Phi \to X$ is a *generalized representative democracy (GRD)* if the following hold.

- If $V = \{v\}$ is a singleton, then $f(\phi) = \phi(v)$.
- If V is not a singleton, there exists a partition $\{V_1, \ldots, V_d\}$ of V into d sets such that

$$f(\phi) = \mathrm{m}(f_1(\phi|_{V_1}), f_2(\phi|_{V_2}), \ldots, f_d(\phi|_{V_d})),$$

  where each $f_i\colon X^{V_i} \to X$ is some generalized representative democracy rule, $\phi|_{V_i}$ is ϕ restricted to V_i, and m is the majority rule.

Any GRD rule is associated with a rooted tree that captures voters' hierarchical structure (as in Figure 3.4 for the case of $d = 3$). A GRD voting rule is equitable if, in the induced tree, the vertices in each level have the same degree.[13]

[13] Intuitively, the permutations required to shift one voter's role into another require the shift of that voter's entire "county" into the target role's "county," which can be done only when their numbers coincide.

The following result characterizes the size of winning coalitions in GRD voting rules.

Theorem 12. *If f is an equitable generalized representative democracy rule for n voters, then a winning coalition must have size at least $n^{\log_3 2}$. Conversely, for arbitrarily large n, there exist equitable generalized representative democracy voting rules with winning coalitions of size $n^{\log_3 2}$.*

There is an intriguing connection between this characterization and the so-called Hausdorff dimension of the Cantor set, which is $\log_3 2 \approx 0.63$.[14] The connection arises from the fact that GRD rules with the smallest winning coalitions are those in which, at each level, the subdivision is into three groups. In such rules, to construct a winning coalition, two of the three top "representatives" need to agree. Then, two of the voters of these representatives need to agree, and so on recursively. This precisely mimics the classical construction of the Cantor set.

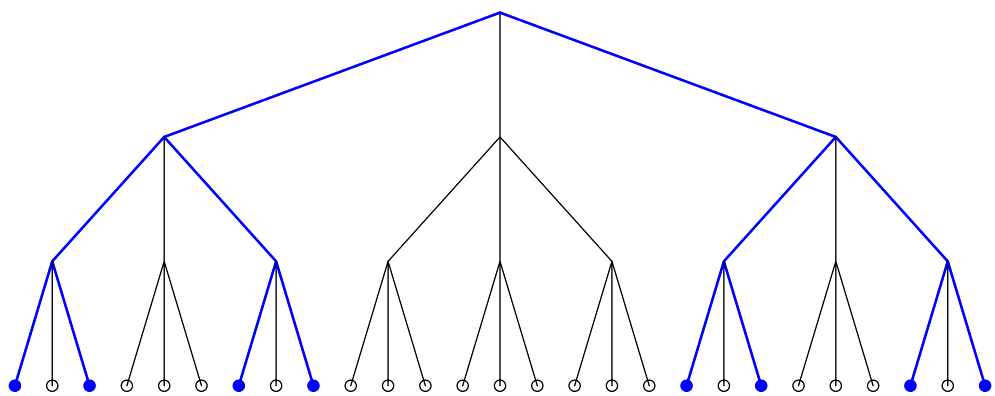

Figure 3.4: Generalized representative democracy voting rule. The leafs of the tree (at the bottom) represent the voters. At each intermediate node the results of the three nodes below are aggregated by majority.

3.4 k-equitable voting rules

So far, we have focused on voting rules in which individuals are indifferent between roles. Naturally, one could extend the notion and contemplate rules that are robust to larger coalitions of voters changing their roles in the population. This section

[14] The Cantor set can be constructed by starting from, say, the unit interval and iteratively deleting the open middle third of any sub-interval remaining. That is, in the first iteration we are left with $[0, 1/3] \cup [2/3, 1]$, in the second iteration we are left with $[0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$, etc. The fractal or Hausdorff dimension is a measure of "roughness" of a set. See Peitgen et al. (1993) and references therein.

analyzes such rules for arbitrary size k of coalitions. With these harsher restrictions on collective-choice procedures, results similar to May's reemerge, although with important caveats.

Definition 2. *A voting rule is k-equitable for $k \geq 1$ if, for every pair of ordered k-tuples $(v_1, \ldots, v_k)$ and $(w_1, \ldots, w_k)$ (with $v_i \neq v_j$ and $w_i \neq w_j$ for all $i \neq j$), there is a permutation $\sigma \in \mathrm{Aut}_f$ such that $\sigma(v_i) = w_i$ for $i = 1, \ldots, k$.*

Intuitively, k-equitable voting rules are ones in which every group of k voters has the same "joint role" in the election. This restriction is certainly harsher than that imposed for equitable rules. Indeed, consider the representative democracy example of Figure 3.1. Suppose Alex is assigned the role of 1, while Bailey is assigned the role of 2.The implication of 2-equity is that the pair (Alex, Bailey) could potentially be associated with any pair (i, j). But this is clearly not true here, since Alex and Bailey are in the same county, and thus it is impossible to associate them with, e.g., the roles of 1 and 4.[15]

In terms of roles and abstract voting rules, k-equity admits the following generalization of Proposition 2A.

Proposition 2B. *An abstract voting rule $f\colon X^R \to X$ is k-equitable iff there is a set of assignments A such that*

1. *$f_a = f_b$ for all $a, b \in A$.*
2. *For each set of k roles $\{r_1, \ldots, r_k\} \subseteq R$ and k voters $\{v_1, \ldots, v_k\} \subset V$ there is an $a \in A$ with $a(v_i) = r_i$ for $i \in \{1, \ldots, k\}$.*

Thus, there is a menu of assignments that all yield the same rule, but that places no restrictions on which roles a coalition of k voters takes. A perhaps illuminating analogy would be to think of equity as corresponding to strategy-proofness, and to think of k-equity as corresponding to strategy-proofness for coalitions of size k. Alternatively, k-equity is reminiscent of group-envy-freeness notions considered in allocation problems (see, e.g., Varian, 1974). In a sense, k is a parameter that interpolates between equity, corresponding to $k = 1$, and May's symmetry, corresponding to $k = n$.

[15] In group-theoretic terms, f is k-equitable if and only if the group Aut_f is *k-transitive*.

We begin by examining 2-equitable rules. Certainly, majority rule is 2-equitable. As can be easily verified, none of the voting rule examples mentioned so far, other than majority, is 2-equitable. As it turns out, for most population sizes, winning coalitions of size at least $n/2$ are endemic to 2-equitable voting rules.

We say that almost every natural number satisfies a property P if the subset $N_P \subseteq \mathbb{N}$ of the natural numbers that have property P satisfies

$$\lim_{n\to\infty} \frac{|N_P \cap \{1,\ldots,n\}|}{n} = 1.$$

Theorem 13. *For almost every natural number n, every 2-equitable voting rule for n voters has no winning coalitions of size less than $n/2$. In particular, for almost all n, the only 2-equitable, neutral, positively responsive voting rule is majority.*

Thus, for almost all n, the assumption of symmetry in May's Theorem can be substituted with the much weaker assumption of 2-equity.

The proof of Theorem 13 relies on modern group-theoretical results that were not available when May's Theorem was introduced Cameron et al. (1982). As it turns out, there is a vanishing share of integers for which there exist 2-transitive groups that are neither the set of all permutations nor the set of even permutations. We show in Lemma 21 in the appendix that those latter groups yield winning coalitions of size at least $n/2$, implying the result.

Theorem 13 states that for most population sizes, 2-equitable rules imply large winning coalitions. This notably does not hold for *all* population sizes. In Appendix 3.9, we construct 2-equitable and 3-equitable voting rules with small winning coalitions, which apply to arbitrarily large population sizes.

When considering more stringent equity restrictions, results are much starker and conclusions hold for *all* population sizes.

Theorem 14. *Every 6-equitable voting rule has no winning coalitions of size less than $n/2$. In particular, the only 6-equitable, neutral, positively responsive voting rule is majority.*[16]

The proof of Theorem 14 relies on discoveries from the 1980's and 1990's that showed that, for any n, the only 6-transitive groups are the group of all permutations

[16]Furthermore, if $n > 24$, the same conclusions follow from the weaker assumption of 4-equity.

and the group of even permutations. These results are a consequence of the successful completion of a large project, involving thousands of papers and hundreds of authors, called the *Classification of Finite Simple Groups*, see Aschbacher et al. (2011).

We stress that k-equity is a strong restriction. Nonetheless, for any fixed k, k-equity is a far weaker restriction than May's symmetry. In that respect, Theorem 14, like Theorem 13, offers a strengthening of May's original result.

3.5 Towards a characterization of equitable voting rules

As our examples throughout illustrate, the set of equitable voting rules is broad and does not admit a simple universal procedural description. Their full characterization would be as complex as the full classification of finite simple groups alluded to above, and hence is beyond the scope of this paper.

We start here by classifying the equitable voting rules for electorate sizes of the form $n = p$ and $n = p^2$ with p a *prime*. We do so for two reasons. First, these cases are easier to handle, while still illustrating some of the complexities entailed in the general characterization of equitable voting rules. To gain some intuition for why these cases are easier, consider the representative democracy rule. When $n = p$, the only representative democracy rule is majority, since there is no way to divide the voters into non-singleton counties of equal size. In particular, the class of equitable rules for $n = p$ is drastically restricted. The second reason we focus on these cases is that it allows us to contemplate general voting rules that can be applied for all electorate sizes n.

For our classification, it will be useful to define *cyclic* voting rules and 2-*cyclic* voting rules.

A voting rule f for n voters is said to be *cyclic* if one can identify the voters with the set $\{0, \ldots, n-1\}$ in such a way that the permutation $\sigma\colon \{0, \ldots, n-1\} \to \{0, \ldots, n-1\}$ given by $\sigma(i) = i + 1 \bmod n$ is an automorphism of f. Intuitively, a voting rule is cyclic if the voters can be arranged on a circle in such a way that rotating, or shifting all voters one space to the right—tantamount to an application of σ—does not affect the outcome. An example of a cyclic voting rule is the longest-run rule described in §3.3. A perhaps less obvious example is that of the representative democracy rule; the arrangement on the cycle entails positioning members of the same county equidistantly along the cycle. In the example of Figure 3.4 in which voters $\{1, 2, \ldots, 9\}$ are arranged in counties $\{\{1, 2, 3\}, \{4, 5, 6, \}, \{7, 8, 9\}\}$, if we arrange the voters along the cycle by the order $(1, 4, 7, 2, 5, 8, 3, 6, 9)$, then applying σ, results in the order $(4, 7, 2, 5, 8, 3, 6, 9, 1)$, and so the first county is mapped to the second, the second is mapped to the third, and the third back to the first. Thus, the voting rule is unchanged by σ, and hence it is cyclic.

A voting rule f for $n = n_1 \times n_2$ voters is said to be 2-*cyclic* if one can identify the voters with the set $\{0, \ldots, n_1 - 1\} \times \{0, \ldots, n_2 - 1\}$ in such a way that the permutations σ_1 and σ_2 given by

$$\sigma_1(i_1, i_2) = (i_1 + 1 \bmod n_1,\ i_2)$$
$$\sigma_2(i_1, i_2) = (i_1,\ i_2 + 1 \bmod n_2)$$

are automorphisms of f. Intuitively, a voting rule is 2-cyclic if one can arrange the voters on an $n_1 \times n_2$ grid such that shifting all voters to the right (and wrapping the rightmost ones back to the left) or shifting all voters up (again wrapping the topmost ones to the bottom) does not affect the outcome. It is easy to see that the cross committee consensus rule is an example of such a rule, as is the representative democracy rule.

Proposition 3. *Let $f\colon X^V \to X$ be an equitable voting rule, and let p be prime. If $n = p$, then f is cyclic. If $n = p^2$, then f is either cyclic, or 2-cyclic, or both.*

The generalization to $n = p^d$ for $d > 2$ is more intricate, and is not a straightforward extension to higher dimensional d-cyclic rules. See our discussion in §3.8.

This proposition has an important implication to the design of simple, equitable voting rules that are not tailored to particular electorate sizes. Majority rule is one such rule—one only needs to tally the votes and consider the difference between the number of supporters of one alternative relative to the other. The longest-run voting rule is another such example. In fact, any such rule must also work for electorates comprised of a prime number of voters, and in particular *must* be cyclic. It would be interesting to understand whether a far larger class of rules than cyclic voting rules can work for *almost all* electorate sizes.

3.6 Conclusions

In this paper we study equity, a notion of procedural fairness that captures equality between different voters' roles.

The voting rules that satisfy May's symmetry axiom admit a simple description: a voting rule is symmetric, or anonymous, if and only if the outcome depends only on the number of voters who choose each possible vote. In contrast, the set of equitable voting rules is much richer, and its complexity is intimately linked to frontier problems in mathematics; in particular, the classification of finite simple groups. This paper includes a number of diverse examples, including generalized representative democracy, cross committee consensus, and a number of additional constructions that appear in the appendix, but the class of all equitable rules is larger yet. Understanding which equitable rules satisfy different desirable conditions—non-malleability, inclusiveness, etc.—could be an interesting avenue for future research.

We believe the approach taken here could potentially be useful for various other contexts. For example, symmetric games are often thought of as ones in which *any* permutation of players' identities does not affect individual payoffs (e.g., Dasgupta and Maskin, 1986, page 18). As is well known, such finite games have symmetric equilibria. Interestingly, in his original treatise on games, Nash took an approach to symmetry that is similar to ours, studying the automorphism group of the game. He showed that equity, analogously defined for games, suffices for the existence of symmetric equilibria: that is, it suffices that for every two players v and w there is an automorphism of the game that maps v to w (Nash, 1951, page 289).[17] It would be interesting to explore further the consequences of equity so defined in more general strategic interactions.

3.7 A primer on finite groups

This section contains what is essentially a condensed first chapter of a book on finite groups (see, e.g., Rotman, 2012), and is provided for the benefit of readers who are not familiar with the topic. The terms and results covered here suffice to prove the main results of this paper.

Denote by $N = \{1, \ldots, n\}$. A *permutation* of N is a bijection $g \colon N \to N$. The inverse of a permutation g is denoted by g^{-1} (so $g^{-1}(g(i)) = i$), and the composition of two permutations g and h is simply gh; i.e., if $k = gh$ then $k(i) = g(h(i))$.

A *group*—for our purposes—will be a non-empty set of permutations that (1) contains g^{-1} whenever it contains g, and (2) contains gh whenever it contains both

[17] Theorem 2 in Nash (1951) states that "Any finite game has a symmetric equilibrium point." Now, Nash's definition of a symmetric equilibrium is a strategy profile $\mathfrak{s}$ such that $\mathfrak{s}_i = \mathfrak{s}_j$ whenever there is an automorphism χ of the game with $\chi(i) = j$. In particular, for there to be a symmetric strategy in the sense we usually consider, where all players use the same strategy, it suffices for there to be a transitive automorphism group, i.e., one in which for every i, j there is an automorphism χ with $\chi(i) = j$.

g and h. It follows from this definition that every group must include the trivial, identity permutation e that satisfies $e(i) = i$ for all i.

Groups often appear as sets of permutations that *preserve some invariant*. In our case, Aut_f is the group of permutations of the voters that preserves every outcome of f. It is easy to see that Aut_f is indeed a group.

A subgroup H of G is simply a subset of G that is also a group. Given $g \in G$, we denote

$$gH = \{gh \in G \ : \ h \in H\}.$$

The sets gH are in general not subgroups, and are called the *left cosets* of H (the right cosets are of the form Hg). It is easy to verify that all left cosets are disjoint, and that each has the same size as H. It follows that the size of G is divisible by the size of H.

Given an element $i \in N$, we denote by G_i the set of permutations that fix i. That is, $g \in G_i$ if $g(i) = i$. G_i is a subgroup of G. It is called the *stabilizer* of i.

The *G-orbit* of $i \in N$ is the set of $j \in N$ such that $j = g(i)$ for some $g \in G$, and is denoted by Gi. As it turns out, if j is in the orbit of i then the set of $g \in G$ such that $g(i) = j$ is a coset of the stabilizer G_i. It follows that there is a bijection between the orbit Gi and the cosets of G_i. This is called the *Orbit-Stabilizer Theorem.*

Recall that G is transitive if for all i, j there is a $g \in G$ such that $g(i) = j$. This is equivalent to there existing only a single G-orbit, or that j is in the orbit of i for every i, j. Therefore, if G is transitive, the orbit Gi is of size n, and since we can identify this orbit with the cosets of G_i, there are n such cosets. Since they are all the same size as G_i, and since they form a partition of G, each coset of G_i must be of size $|G|/n$. We will use this fact in the proof of Theorem 11.

3.8 Proofs

Proofs of Propositions 1A, 1B, 2A, and 2B

Proof of Proposition 1A. We begin by showing that if f is anonymous, then all assignments are equivalent under f.

Assume that f is symmetric, and let a and b be two assignments. Define $\sigma = b \circ a^{-1}$. Then, by definition of symmetry, for every $\phi \in X^V$,

$$f(\phi \circ a^{-1} \circ \sigma^{-1}) = f((\phi \circ a^{-1})^{\sigma}) = f(\phi \circ a^{-1}).$$

Since $b^{-1} = a^{-1} \circ \sigma^{-1}$, it follows that $f(\phi \circ a^{-1}) = f(\phi \circ b^{-1})$ and hence that $f_a(\phi) = f_b(\phi)$, so a and b are equivalent. Since a and b were arbitrary, it follows that all assignments are equivalent under f.

Conversely, suppose that all assignments are equivalent under f. We need to show that for any $\sigma \in S(R)$ and any $\phi \in X^R$, $f(\phi) = f(\phi^\sigma)$. Fix σ and ϕ, let a be any assignment, and let $b = \sigma \circ a$. Then since a and b are equivalent,

$$f(\phi) = f((\phi \circ a) \circ a^{-1}) = f_a(\phi \circ a) = f_b(\phi \circ a) = f(\phi \circ (a \circ b^{-1})) = f(\phi \circ \sigma^{-1}) = f(\phi^\sigma).$$

Since σ and ϕ were arbitrary, it follows that f is symmetric. □

In analogy with the notion of equivalence of roles given in Section 3.2, say that two k-tuples of roles $\mathbf{r}$ and $\mathbf{s}$ are equivalent if there is a k-tuple $\mathbf{v}$ of voters and a pair of assignments a and b such that $f_a = f_b$, $a(\mathbf{v}) = \mathbf{r}$, and $b(\mathbf{v}) = \mathbf{s}$. Proposition 1B then follows from the next proposition by setting k to 1:

Proposition 3. *An abstract voting rule $f\colon X^R \longrightarrow X$ is k-equitable if and only if all k-tuples of roles are equivalent under f.*

Proof. Fix an assignment a. The rule f is k-equitable if and only if f_a is k-equitable, which holds if and only if for every pair of k-tuples $(v_1, \dots, v_k)$ and $(w_1, \dots, w_k)$ of voters there is a $\sigma \in S(V)$ such that $\sigma((v_1, \dots, v_k)) = (w_1, \dots, w_k)$ and for all $\phi \in X^V$, $f_a(\phi) = f_a(\phi^\sigma)$. Since

$$\begin{aligned} f_a(\phi^\sigma) &= f((\phi \circ \sigma^{-1}) \circ a^{-1}) \\ &= f(\phi \circ (a \circ \sigma)^{-1}) \\ &= f_{a\circ\sigma}(\phi), \end{aligned}$$

it follows that f is k-equitable if and only if for every pair of k-tuples $(v_1, \dots, v_k)$ and $(w_1, \dots, w_k)$ of voters there is a $\sigma \in S(V)$ such that $\sigma((v_1, \dots, v_k)) = (w_1, \dots, w_k)$ and $f_a = f_{a\circ\sigma}$.

Now, this holds if and only if for every pair of k-tuples $(r_1, \dots, r_k)$ and $(s_1, \dots, s_k)$ of roles there is a $\sigma \in S(V)$ such that $\sigma(a^{-1}(r_1, \dots, r_k)) = a^{-1}(s_1, \dots, s_k)$ and $f_a = f_{a\circ\sigma}$. But this holds if and only if for every such pair of k-tuples of roles, there is a $\sigma \in S(V)$ and a k-tuple of voters $(v_1, \dots, v_k)$ such that $f_a = f_{a\circ\sigma}$, $a((v_1, \dots, v_k)) = (r_1, \dots, r_k)$, $(a \circ \sigma)((v_1, \dots, v_k)) = (s_1, \dots, s_k)$, which holds if and only if every pair of k-tuples of roles is equivalent under f. □

We now prove Proposition 2B. Proposition 2A then follows directly by setting k to 1.

Proof of Proposition 2B. We first show that if f is k-equitable, then there is a set A as above.

Assume f is k-equitable. Fix an assignment a, and let

$$A = \{b \text{ an assignment s.t. } f_a = f_b\}.$$

(1) is immediate from the definition of A. To see that (2) holds, note that for any k-tuple of roles $(r_1, \ldots, r_k)$ and any k-tuple of voters $(v_1, \ldots, v_k)$, it follows from a result analogous to Proposition 3 that since f is k-equitable, there are assignments c and d such that $c((v_1, \ldots, v_k)) = (r_1, \ldots, r_k)$, $d((v_1, \ldots, v_k)) = a((v_1, \ldots, v_k))$, and $f_c = f_d$. Now, $f_c = f_d$ implies that for all $\phi \in X^V$, $f_c(\phi) = f_d(\phi)$, which implies that for all $\phi \in X^V$, $f_c(\phi \circ (a^{-1} \circ d)) = f_d(\phi \circ (a^{-1} \circ d))$. It follows that, for all $\phi \in X^V$,

$$\begin{aligned}
f_{c \circ d^{-1} \circ a}(\phi) &= f(\phi \circ (c \circ d^{-1} \circ a)^{-1}) \\
&= f((\phi \circ (a^{-1} \circ d)) \circ c^{-1}) \\
&= f_c(\phi \circ (a^{-1} \circ d)) \\
&= f_d(\phi \circ (a^{-1} \circ d)) \\
&= f((\phi \circ (a^{-1} \circ d)) \circ d^{-1}) \\
&= f(\phi \circ a^{-1}) \\
&= f_a(\phi),
\end{aligned}$$

and hence $f_{c \circ d^{-1} \circ a} = f_a$. But this implies that $c \circ d^{-1} \circ a \in A$. Since $(c \circ d^{-1} \circ a)((v_1, \ldots, v_k)) = c(d^{-1}(a((v_1, \ldots, v_k)))) = c((v_1, \ldots, v_k)) = (r_1, \ldots, r_k)$, (2) then follows.

We now show that if there is a set A as above, then f is k-equitable.

Assume there is such an A. By a result analogous to Proposition 3, it is sufficient to show that all k-tuples of roles are equivalent under f. But this follows immediately from (2) and the definition of equivalence of k-tuples of roles. □

Proof of Lemma 19

Proof of Lemma 19. Let f be the voting rule defined as follows. For a voting profile ϕ, if there is a set $W \in \mathcal{W}$ such that $\phi(w) = 1$ for all $w \in W$, then $f(\phi) = 1$, and similarly, if there is a set $W \in \mathcal{W}$ such that $\phi(w) = -1$ for all $w \in W$, then $f(\phi) = -1$. This is well-defined, since if there are two such sets W, they must agree because they intersect. If there are no such sets, then $f(\phi)$ is determined by majority.

That f is neutral follows immediately from the symmetry in the definition of f when some $W \in \mathcal{W}$ agrees on either 1 or -1 and the fact that majority is neutral. To see that f is positively responsive, suppose that $f(\phi) \in \{0, 1\}$, $\phi'(x) \geq \phi(x)$ for all $x \in V$, and $\phi'(y) > \phi(y)$ for some $y \in V$. Since $f(\phi) \neq -1$, there is no set $W \in \mathcal{W}$ such that $\phi(x) = -1$ for all $x \in W$, hence the same is true for ϕ'. If there is some set $W \in \mathcal{W}$ such that $\phi'(x) = 1$ for all $x \in W$, then $f(\phi') = 1$. If not, then the same is true of ϕ, and hence by positive responsiveness of majority, $f(\phi') = 1$.

Finally, it is immediate from the definition of f that every $W \in \mathcal{W}$ is a winning coalition. □

Proof of Theorem 10

Proof of Theorem 10. The longest-run voting rule is equitable, since any rotation of the cycle is an automorphism. That is, for every k, the map $\sigma \colon V \to V$ defined by $\sigma(i) = i + k \bmod n$ leaves the outcome unchanged. Furthermore, for every pair of voters i, j, if we set $k = i - j$, then $\sigma(i) = j$.

For every V of size n, we claim that the longest-run rule $\ell \colon X^V \to X$ has winning coalitions of size $2\lceil\sqrt{n}\,\rceil - 1$.

Let

$$W = \{0, \ldots, \lceil\sqrt{n}\,\rceil - 1\} \cup \{w \,:\, w \bmod \lceil\sqrt{n}\,\rceil = 0\}.$$

Any run that is disjoint from W is of length at most $\lceil\sqrt{n}\,\rceil - 1$ since the second set in the definition of W is comprised of voters who are at most $\lceil\sqrt{n}\,\rceil$ apart. However, the first set is a contiguous block of length $\lceil\sqrt{n}\,\rceil$. Hence, if all $w \in W$ vote identically in $\{-1, 1\}$, the longest run will be a subset of w, and hence the outcome will be the vote cast by all members of W. It then follows that W is a winning coalition.

Finally,

$$\begin{aligned}
|W| &= |\{0, \ldots, \lceil\sqrt{n}\,\rceil - 1\}| + |\{w \,:\, w \bmod \lceil\sqrt{n}\,\rceil = 0\}| - |\{0\}| \\
&= \lceil\sqrt{n}\,\rceil + |\{w \,:\, w \bmod \lceil\sqrt{n}\,\rceil = 0\}| - 1 \\
&\leq 2\lceil\sqrt{n}\,\rceil - 1.
\end{aligned}$$

Since every superset of a winning coalition is again a winning coalition, the result follows. □

Proof of Theorem 11

Readers who are not familiar with the theory of finite groups are encouraged to read §3.7 before reading this proof.

Recall that the group of all permutations of a set of size n is denoted by S_n.

The next lemma shows that if a group G acts transitively on $\{1, \ldots, n\}$, then any set S that intersects all of its translates (i.e., sets of the form gS for $g \in G$) must be of size at least $\sqrt{n}$. The proof of the theorem will apply this lemma to a winning coalition S.

Lemma 20. *Let $G \subset S_n$ be transitive, and suppose that $S \subseteq V$ is such that for all $g \in G$, $gS \cap S \neq \emptyset$. Then $|S| \geq \sqrt{n}$.*

Proof. For any $v, w \in V$, define $\Gamma_{v,w} = \{g \in G : g(v) = w\}$. Then $\Gamma_{v,w}$ is a left coset of the stabilizer of v. Hence, and since the action is transitive, it follows from the Orbit-Stabilizer Theorem that $|\Gamma_{v,w}| = \frac{|G|}{n}$. If $gS \cap S \neq \emptyset$ for all $g \in G$, then for any $g \in G$, there exists $v, w \in S$ such that $g(v) = w$, hence

$$\bigcup_{v,w \in S} \Gamma_{v,w} = G.$$

So

$$|G| = \left| \bigcup_{v,w} \Gamma_{v,w} \right| \leq \sum_{v,w} |\Gamma_{v,w}| = |S|^2 \frac{|G|}{n},$$

and we conclude that $|S| \geq \sqrt{n}$. □

Our lower bound (Theorem 11) is an immediate corollary of this claim.

Proof of Theorem 11. Let f be an equitable voting rule for the voter set V. Suppose that $W \subseteq V$ is a winning coalition for f. Then, for every $\sigma \in \mathrm{Aut}_f$, it must be the case that $\sigma(W) \cap W \neq \emptyset$ (otherwise, f would not be well-defined). Hence, it follows from Lemma 20 that $|W| \geq \sqrt{n}$. □

Proof of Theorem 12

Proof of Theorem 12. Define $C(n)$ to be the smallest size of any winning coalition in any generalized representative democracy rule for n voters. We want to show that $C(n) \geq n^{\log_3 2}$.

If $n = 1$, a winning coalition must be of size 1, which is $\geq 1^{\log_3 2}$.

If $n > 1$, any generalized voting rule f is of the form $f(\phi) = \mathrm{m}(f_1(\phi|_{V_1}), f_2(\phi|_{V_2}), \ldots, f_d(\phi|_{V_d}))$. Because the voting rule is equitable, the functions f_i are all isomorphic, and so have minimal winning coalitions of the same size. A minimal winning coalition for f would then need to include a strict majority of these, which is of size at least $\frac{d+1}{2}$. Therefore,[18]

$$C(n) \geq \min_{d|n} \frac{d+1}{2} \cdot C(n/d). \tag{3.1}$$

Assume by induction that $C(m) \geq m^{\log_3 2}$ for all $m < n$. Then for $d|n$,

$$\frac{d+1}{2} \cdot C(n/d) \geq \frac{d+1}{2} \cdot \left(\frac{n}{d}\right)^{\log_3 2} = n^{\log_3 2} \cdot \frac{d+1}{2} \cdot d^{-\log_3 2}.$$

Denote $h(d) = \frac{d+1}{2} \cdot d^{-\log_3 2}$, so that

$$\frac{d+1}{2} C(n/d) \geq n^{\log_3 2} h(d).$$

Note that $h(d) \geq 1$. To see this, observe that $h(3) = 1$, and

$$h'(d) = \frac{d^{-\log 6/\log 3}(d \log \frac{3}{2} - \log 2)}{2 \log 3} > 0$$

for $d \geq 3$, and so $h(d) \geq 1$ for $d \geq 3$.

We have thus shown that

$$\frac{d+1}{2} C(n/d) \geq n^{\log_3 2},$$

and so by (3.1), $C(n) \geq n^{\log_3 2}$.

To see that $C(n) = n^{\log_3 2}$ for arbitrarily large n, consider the following GRD rule (see Figure 3.4). Take n to be a power of 3, and let f be defined recursively by partitioning at each level into three sets $\{V_1, V_2, V_3\}$ of equal size. A simple calculation shows that the winning coalition recursively consisting of the winning coalitions of any two of $\{V_1, V_2, V_3\}$ (e.g., V_1 and V_3, as in Figure 3.4), is of size $n^{\log_3 2}$. □

The construction of small winning coalitions in the last part of the proof mimics the construction of the Cantor set.

[18] Here and below $d|n$ denotes that d is a divisor of n.

Proof of Theorems 13 and 14

The group of all even permutations is called the *alternating group* and is denoted A_n.

Lemma 21. *Let f be a voting rule for n voters. If Aut_f is either S_n or A_n, then every winning coalition for f has size at least $n/2$.*

Proof. Suppose $W \subseteq V$ is a winning coalition for f with $|W| = k < n/2$. Label the voters V with labels $1, \ldots, n$ such that $W = \{1, \ldots, k\}$, and let π be the permutation of V given by $\pi(i) = n + 1 - i$ for $i = 1, \ldots, n$. If $\lfloor n/2 \rfloor$ is odd, let π be the map above composed with the map that exchanges 1 and 2. It follows that π is in the alternating group, and hence $\pi \in \mathrm{Aut}_f$. However, $\pi(W) \cap W = \emptyset$ since $k < n+1-k$, contradicting the assumption W is a winning coalition. □

Proof of Theorem 13. Denote by $\eta(n)$ the number of positive integers $m \leq n$ for which there is no 2-transitive group action on a set of m elements except for S_m and A_m. It follows from the main theorem in Cameron et al. (1982) that $n - \eta(n)$ is at most $3n/\log(n)$ for all n large enough. Since

$$\lim_{n\to\infty} \frac{3n/\log(n)}{n} = 0,$$

it follows that

$$\lim_{n\to\infty} \frac{\eta(n)}{n} = \lim_{n\to\infty} 1 - \frac{n - \eta(n)}{n} = 1,$$

and so the claim follows from Lemma 21. □

Proof of Theorem 14. The only 4- or 5-transitive finite groups aside from the alternating and symmetric groups are the Mathieu groups, with the largest action on a set of size 24 (Dixon and Mortimer, 1996). Hence, for $n > 24$, every 4- or 5-transitive voting rule must have either S_n or A_n as an automorphism group. Furthermore, the *only* 6-transitive groups are S_n or A_n (again, see Dixon and Mortimer, 1996). Hence, the result follows immediately from Lemma 21. □

Proof of Proposition 3

This proposition's proof follows directly from group theory results that are classical, but that are not covered in our primer in §3.7, and which we now review briefly.

Let G be a group. The *order* of G is simply its size. The order of $g \in G$ is the smallest n such that g^n is the identity. Given a prime p, we say that a group P is a

p-group if the orders of all of its elements are equal to powers of p. We assume for the remainder of this section that p is prime.

Let G be a finite group with $|G| = p^d \cdot m$, where $d \geq 1$, and m is not divisible by p. Sylow (1872) proved that, in this case, G has a subgroup P that is a p-group of order p^d. Such groups are called Sylow p-groups in his honor. The following lemma states an important and well-known fact (see, e.g., Wielandt, 2014, Theorem 3.4') regarding Sylow p-groups.

Lemma 22. *Let G act transitively on a set V of size $n = p^d$ for some $d \geq 1$. Any Sylow p-subgroup of G acts transitively on V.*

Proof. Since the order of G is divisible by the size of V, $|G| = p^{d+\ell} \cdot m$ for some $\ell \geq 0$ and m not divisible by p. Let P be a Sylow p-subgroup, so that $|P| = p^{d+\ell}$.

For any $i \in V$, the size of the P-orbit Pi divides $|P| = p^{d+\ell}$, and so is equal to p^{d-a} for some $a \geq 0$. Now, the size of the stabilizers P_i and G_i is $|P_i| = |P|/|Pi| = p^{\ell+a}$ and $|G_i| = p^\ell \cdot m$. Since P_i is a subgroup of G_i, $|P_i|$ divides $|G_i|$, and so $a = 0$, $|Pi| = p^d = n$, and P acts transitively on V. □

The *center* of a group is the collection of all of its elements that commute with all the group elements: $\{g \in G \ : \ gh = hg \text{ for all } h \in G\}$. This is easily seen to also be a subgroup of G.

Lemma 23. *Every non-trivial p-group has a non-trivial center.*

For a proof, see Theorem 6.5 in Lang (2002). Here and below, a non-trivial group is a group of order larger than 1.

Let G be a group that acts transitively on a set V, and let Z be the center of G. Denote by G/Z the set of left cosets of Z, and let $\hat{V}$ be the set of Z-orbits of V. If $v, w \in V$ are in the same Z-orbit, then $g(v)$ and $g(w)$ are also in the same Z-orbit, since if $z(v) = w$ then $z(g(v)) = g(z(v)) = g(w)$. Hence, each $g \in G$ induces a permutation on $\hat{V}$. Note that $g, h \in G$ induce the same permutation on $\hat{V}$ if they are in the same element of G/Z. Hence, we can think of G/Z as a group of permutations of $\hat{V}$. This group must act transitively on $\hat{V}$ since G acts transitively on V. Furthermore, if a subgroup of G/Z acts transitively on $\hat{V}$, then the union of the cosets it includes is a subgroup of G that acts transitively on V.

Lemma 24. *Let G be a group acting transitively on a set V of size p^d, for some $d \geq 1$. There exists a p-subgroup R of G of size p^d acting transitively on V with trivial stabilizers.*

Proof. Let P denote a Sylow p-subgroup of G. By Lemma 22, the action of P on V is also transitive.

Let Z denote the center of P. Since P is non-trivial, by Lemma 23, Z is non-trivial. We claim that the Z action on V has trivial stabilizers. To see this, assume that $h(v) = v$ for some v, and choose any $w \in V$. Since P acts transitively, there is some $g \in P$ such that $g(v) = w$. Since h commutes with g,

$$h(w) = h(g(v)) = g(h(v)) = g(v) = w,$$

and so, since w was arbitrary, h is the identity. Hence, $Z_v = \{e\}$ for every $v \in V$. Note that by the Orbit-Stabilizer Theorem, this implies that each Z orbit is equal in size to Z.

If the action of Z is also transitive, we are done, since we can take $R = Z$.

Finally, consider the case that Z does not act transitively.

In this case $P' = P/Z$ acts transitively on $\hat{V}$, the set of the Z-orbits of V. By induction, P' has a subgroup Z' which acts transitively with trivial stabilizers on $\hat{V}$, and hence has size $|\hat{V}|$. Note that since the action of Z has trivial stabilizers, every Z-orbit has size $|Z|$, and so $|\hat{V}| = |V|/|Z| = p^d/|Z|$. Thus, taking R to be the union of the cosets in P/Z that comprise Z', R acts transitively on V. Finally, since this subgroup has size $|Z'| \cdot |Z| = p^d = |V|$, it follows that the action of R has trivial stabilizers.

□

A group is said to be *abelian* if all of its elements commute: $gh = hg$ for all $g, h \in G$. Note that the center of every group is abelian by definition. The structure of abelian groups is simple and well understood: Kronecker's Theorem (Kronecker, 1870, Stillwell, 2012, Theorem 5.2.2) states that every abelian group is a product of cycles of prime powers. That is, if G is an abelian group of permutations of a set V—and assuming without loss of generality that no element of V is fixed by all elements of G—then there is a way to identify V with $\prod_{i=1}^{m}\{0, \ldots, n_i - 1\}$, with each n_i a prime power, so that G is generated by permutations[19] of the form

$$\sigma_k(i_1, \dots, i_m) = (i_1, \dots, i_{k-1}, i_k + 1 \bmod n_k, i_{k+1}, \dots, i_m).$$

As is also well known, every group of order p or p^2 is abelian (Netto, 1892, page 148). From these facts follows the next lemma.

Lemma 25. *Let R be a group of order p or p^2, acting transitively on a set V. In the former case, we can identify V with $\{0, \dots, p-1\}$ so that R is generated by*

$$\sigma(i) = i + 1 \; mod \; p.$$

In the latter case, we can either identify V with $\{0, \dots, p^2-1\}$ so that R is generated by

$$\sigma(i) = i + 1 \; mod \; p^2,$$

or else we can identify V with $\{0, \dots, p-1\}^2$, so that R is generated by

$$\sigma_1(i_1, i_2) = (i_1 + 1 \; mod \; p, i_2)$$
$$\sigma_1(i_1, i_2) = (i_1, i_2 + 1 \; mod \; p).$$

Hence, if R is a subgroup of the automorphism group of a voting rule f, then this rule is cyclic if $n = p$, and is either cyclic, 2-cyclic or both if $n = p^2$.

Proof of Proposition 3. By Lemma 24, Aut_f has a p-subgroup R of order n that acts transitively on V. The claim now follows immediately from Lemma 25. □

When $n = p^d$, with $d > 2$, this proof fails since the group R is no longer necessarily abelian. Non-abelian groups do not have cyclic structure, and thus voting rules for which this group R is not abelian will not be cyclic, 2-cyclic, or higher-dimensional cyclic. We conjecture that such equitable voting rules do indeed exist.

3.9 2-equitable and 3-equitable rules

In this section we construct 2-equitable and 3-equitable rules with small winning coalitions that apply to arbitrarily large population sizes. This construction is rather technically involved and uses finite vector spaces. To glean some intuition, we first explain an analogous construction using standard vector spaces and assuming a continuum of voters.

[19] A group is said to be generated by a set S of permutations if it includes precisely those permutations that can be constructed by composing permutations in S.

Suppose voters are identified with the set of one-dimensional subspaces of $\mathbb{R}^3$: i.e., each voter is identified with a line that passes through the origin. Now suppose winning coalitions are the two-dimensional subspaces: if all voters on a plane agree, that is the election outcome, otherwise the election is undecided.[20] Clearly, the winning coalitions are much smaller than the electorate (or indeed of "half of the voters") in the sense that they have a smaller dimension.

Invertible linear transformations of $\mathbb{R}^3$ permute the one-dimensional subspaces, and the two-dimensional subspaces, and so constitute automorphisms of this voting rule. Equity follows since for any two non-zero vectors v and u, we can find some invertible linear transformation that maps v to u. Moreover, the voting rule is also 2-equitable—given a pair of distinct voters (v_1, v_2), and given another such pair (u_1, u_2), we can find some invertible linear transformation that maps the former to the latter. Thus, *every pair* of voters plays the same role.

In Theorem 15 below we construct 2-equitable voting rules for finite sets of voters, using finite vector spaces instead of $\mathbb{R}^3$. Figure 3.5 shows a 2-equitable voting rule constructed in this way, for 7 voters. In the figure, every three co-linear nodes form a winning coalition, as well as the three nodes on the circle.[21] In this construction, the size of the winning coalition is exactly $\sqrt{n}$ (rounded up to the nearest integer), which matches the lower bound of $\sqrt{n}$ in Theorem 11.

Theorem 15. *Let the set of voters be of size $n = q^2 + q + 1$, for prime q. Then there is a 2-equitable voting rule with a winning coalition of size exactly equal to $\sqrt{n}$, rounded up to the nearest integer.*

More generally, a similar statement holds when $n = q^2 + q + 1$ and $q = p^k$ for some $k \geq 1$ and p prime. The example in Figure 3.5 corresponds to the case $q = 2$.

Proof. Let $\mathbb{F}_q$ denote the finite field with q elements.[22]

Given a positive integer m, $\mathbb{F}_q^m$ is a vector space, where the scalars take values in $\mathbb{F}_q$: it satisfies all the axioms that (say) $\mathbb{R}^3$ satisfies, but for scalars that are in $\mathbb{F}_q$ instead

[20] This rule is well defined since every pair of two-dimensional subspaces intersects, so no two winning coalitions are disjoint.

[21] Figure 3.5 depicts what is commonly referred to as a Fano plane in finite geometry. It is the finite projective plane of order 2.

[22] $\mathbb{F}_q$ is the set $\{0, 1, \ldots, q-1\}$, equipped with the operations of addition and multiplication modulo q. The primality of q is required to make multiplication invertible.

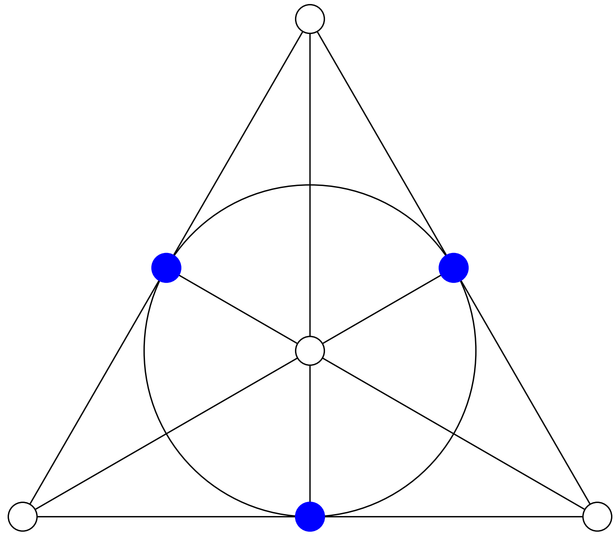

Figure 3.5: Every three co-linear points form a winning coalition, as well as the three points on the circle (marked in blue). This voting rule is 2-equitable.

of $\mathbb{R}$. Indeed, much of the standard theory of linear algebra of $\mathbb{R}^m$ applies in this finite setting, and we will make use of it here.

In particular, we will make use of $GL(m,q)$, the group of invertible, $m \times m$ matrices with entries in $\mathbb{F}_q$. Here, again, the product of two matrices is calculated as usual, but addition and multiplication are taken modulo q. Since $\mathbb{F}_q^m$ is finite, each matrix in $GL(m,q)$ corresponds to a permutation of $\mathbb{F}_q^m$. As in the case of matrix multiplication on $\mathbb{R}^m$, these permutations preserve the 1-dimensional and 2-dimensional subspaces. Moreover, this group acts 2-transitively on the 1-dimensional subspaces, as any two non-colinear vectors (u,v) can be completed to a basis of $\mathbb{F}_q^m$, and likewise starting from (u',v'); then any basis can be carried by an invertible matrix to any other basis.

With this established, we are ready to identify our set of voters with the set of 1-dimensional subspaces of $\mathbb{F}_q^3$. For each 2-dimensional subspace U of $\mathbb{F}_q^3$, define the set S_U of 1-dimensional subspaces (i.e., voters) contained in U. Let $\mathcal{W}$ be the collection of all such sets S_U, and define, using Lemma 19, a voting rule f in which the sets S_U are winning coalitions. We need to verify that any two winning coalitions S_U and $S_{U'}$ are non-disjoint. This simply follows from the fact that every pair of 2-dimensional subspaces intersects in some 1-dimensional subspace, and so it follows that each pair of such winning coalitions will have exactly one voter in common.[23]

A simple calculation shows that the winning coalitions are of size $q + 1$. Since $\sqrt{n} \leq q + 1 \leq \sqrt{n} + 1$, the claim follows. □

To construct 3-equitable rules we will need the following lemma. It allows us to show, using the probabilistic method, that for small automorphism groups we can construct voting rules with small winning coalitions. This is useful for proving that there exist 3-equitable voting rules with small winning coalitions.

Lemma 26. *Let G be a group of m permutations of $\{1, \ldots, n\}$. Then there is a neutral and positively responsive voting rule f such that G is a subgroup of Aut_f, and f has winning coalitions of size at most $2\sqrt{n}\log m + 2$.*

We use this lemma to prove our theorem illustrating the existence of 3-equitable rules with small winning coalitions for arbitrarily large voter populations. We then return to prove the lemma.

Theorem 16. *For n such that $n - 1$ is a prime power, there is a 3-equitable voting rule with a winning coalition of size at most $6\sqrt{n}\log n$.*

Proof. For n such that $n - 1$ is the power of some prime there is a 3-transitive group of permutations of $\{1, \ldots, n\}$ that is of size $m < n^3$.[24] Hence, by Lemma 26, there is a 3-equitable voting rule for n (i.e., a rule with a 3-transitive automorphism group) with a winning coalition of size at most $2\sqrt{n}\log(n^3) = 6\sqrt{n}\log n$. □

It is natural to conjecture that this probabilistic construction is not optimal, and that there exist 3-equitable rules with winning coalitions of size $O(\sqrt{n})$.

The heart of Lemma 26 is the following group-theoretic claim, which states that when G is small then we can find a small set S such that gS and S are non-disjoint for every $g \in G$. These sets gS will be the winning coalitions used to prove Lemma 26. The proof of this proposition uses the *probabilistic method*: we choose S at random from some distribution, and show that, with positive probability, it has the desired property. This proves that there exists a deterministic S with the desired property.

[23] This is the reason that the winning coalitions of this rule are so small and proving a tight match to the lower bound.

[24] The group $PGL(2, n-1)$ acts 3-transitively on the projective line over the field $\mathbb{F}_{n-1}$, and is of size $n(n-1)(n-2) < n^3$.

Proposition 4. *Let a group G of $m > 2$ permutations of $\{1, \ldots, n\}$. Then there exists a set $S \subseteq \{1, \ldots, n\}$ with $|S| \leq 2\sqrt{n}\log m + 2$ such that $\forall g \in G$ we have $gS \cap S \neq \emptyset$.*

Proof. To prove this, we will choose S at random, and prove that it has the desired properties with positive probability. Let $\ell = \lceil \sqrt{n}\log|G| \rceil$. Let $S = S_1 \cup S_2$, where S_1 is any subset of X of size ℓ, and S_2 is the union of ℓ elements of X, chosen independently from the uniform distribution. Hence S includes at most $2\ell \leq 2\sqrt{n}\log|G| + 2$ elements.

We now show that $\mathbb{P}(\forall g \in G : gS \cap S \neq \emptyset) > 0$, and hence there is some set S with the desired property. Note that for any particular $g \in G$, the distribution of gS_2 is identical to the distribution of S_2. Hence

$$\begin{aligned}
\mathbb{P}(gS \cap S = \emptyset) &\leq \mathbb{P}(gS_2 \cap S_1 = \emptyset) \\
&= \mathbb{P}(S_2 \cap S_1 = \emptyset) \\
&= \left(\frac{n-\ell}{n}\right)^{\ell} \\
&\leq e^{-\ell^2/n} \\
&\leq e^{-(\log m)^2}.
\end{aligned}$$

Thus, the probability that there is some $g \in G$ for which $gS \cap S = \emptyset$ is, by taking a union bound, at most

$$me^{-(\log|G|)^2},$$

which is strictly less than 1 for $m > 2$. □

We are finally ready to prove Lemma 26.

Proof of Lemma 26. Let S be the subset of $\{1, \ldots, n\}$ given by Proposition 4. Let $\mathcal{W}$ be the collection of sets of the form gS, where $g \in G$. This is a collection of pairwise non-disjoint sets, since if gS and hS intersect then so do $h^{-1}gS$ and S, which is impossible by the defining property of S. Since $|S| = 2\sqrt{n}\log m + 2$ the claim follows from Lemma 19. □

Bibliography

Daron Acemoglu, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

Itai Arieli and Manuel Mueller-Frank. Multidimensional social learning. *The Review of Economic Studies*, 86(3):913–940, 2019.

Michael Aschbacher, Richard Lyons, Stephen D Smith, and Ronald Solomon. *The Classification of finite simple groups: Groups of characteristic 2 type*, volume 172. American Mathematical Soc., 2011.

Venkatesh Bala and Sanjeev Goyal. Learning from neighbours. *The review of economic studies*, 65(3):595–621, 1998.

Abhijit V Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817, 1992.

Aadyot Bhatnagar. Voting Rules that are Unbiased but not Transitive-Symmetric. *The Electronic Journal of Combinatorics*, 27, 2020.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992.

Peter J. Cameron, Peter M. Neumann, and David N. Teague. On the degrees of primitive permutation groups. *Mathematische Zeitschrift*, 180(3):141–149, 1982. ISSN 1432-1823.

Estelle Cantillon and Antonio Rangel. A graphical analysis of some basic results in social choice. *Social Choice and Welfare*, 19(3):587–611, 2002.

Christophe Chamley. Delays and equilibria with large and small information in social learning. *European Economic Review*, 48(3):477–501, 2004a.

Christophe Chamley. *Rational herds: Economic models of social learning*. Cambridge University Press, 2004b.

Partha Dasgupta and Eric Maskin. The existence of equilibrium in discontinuous economic games, i: Theory. *Review of Economic Studies*, 53(1):1–26, 1986.

John D Dixon and Brian Mortimer. *Permutation Groups*, volume 163. Springer Science & Business Media, 1996.

Pradeep Dubey and Lloyd S Shapley. Mathematical properties of the banzhaf power index. *Mathematics of Operations Research*, 4(2):99–131, 1979.

Darrell Duffie and Gustavo Manso. Information percolation in large markets. *The American Economic Review*, pages 203–209, 2007.

Darrell Duffie, Semyon Malamud, and Gustavo Manso. Information percolation with equilibrium search dynamics. *Econometrica*, 77(5):1513–1574, 2009.

Mark Fey. May's theorem with an infinite population. *Social Choice and Welfare*, 23(2):275–293, 2004.

Benjamin Golub and Evan Sadler. Learning in social networks. *Available at SSRN 2919146*, 2017.

Robert Goodin and Christian List. A conditional defense of plurality rule: Generalizing may's theorem in a restricted informational environment. *American Journal of Political Science*, 50(4):940–949, 2006.

Steven R Grenadier. Information revelation through option exercise. *The Review of Financial Studies*, 12(1):95–129, 1999.

Wade Hann-Caruthers. The role of networks structure in the diffusion of investment decisions. In preparation, 2022.

John R Isbell. Homogeneous games. II. *Proceedings of the American Mathematical Society*, 11(2):159–161, 1960.

Leopold Kronecker. *Auseinandersetzung einiger Eigenschaften der Klassenzahl idealer complexer Zahlen*. Monatsberichte der Akademie der Wiss. zu Berlin, 1870.

Serge Lang. *Algebra*. Springer New York, 2002.

Ilan Lobel and Evan Sadler. Information diffusion in networks through social learning. *Theoretical Economics*, 10(3):807–851, 2015.

Ilan Lobel, Daron Acemoglu, Munther Dahleh, and Asuman Ozdaglar. Rate of convergence of learning in social networks. In *Proceedings of the American Control Conference*, 2009.

Kenneth O. May. A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, 20(4):680–684, 1952.

Anthony J. McGann, Charles A. Smith, Michael Latner, and Alex Keena. *Gerrymandering in America: The House of Representatives, the Supreme Court, and the Future of Popular Sovereignty*. Cambridge University Press, 2016.

Pooya Molavi, Alireza Tahbaz-Salehi, and Ali Jadbabaie. A theory of non-bayesian social learning. *Econometrica*, 86(2):445–490, 2018.

Elchanan Mossel and Ryan O'Donnell. Recursive reconstruction on periodic trees. *Random Structures & Algorithms*, 13(1):81–97, 1998.

Elchanan Mossel, Allan Sly, and Omer Tamuz. Strategic learning and the topology of social networks. *Econometrica*, 83(5):1755–1794, 2015.

John Nash. Non-cooperative games. *Annals of Mathematics*, pages 286–295, 1951.

Eugen Netto. *The theory of substitutions and its applications to algebra*. Inland Press, 1892. Available online at `https://archive.org/details/theoryofsubstitu00nett/page/148`.

Edward W Packel. Transitive permutation groups and equipotent voting rules. *Mathematical Social Sciences*, 1(1):93–100, 1980.

Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe. *Chaos and Fractals: New Frontiers of Science*. Springer, 1993.

Lionel Penrose. The elementary statistics of majority voting. *Journal of the Royal Statistical Society*, 109:53–57, 1946.

William Reiker. *The Theory of Political Coalitions*. New Haven and London: Yale University Press, 1962.

Dinah Rosenberg and Nicolas Vieille. On the efficiency of social learning. *private communication*, 2017.

Joseph J Rotman. *An Introduction to the Theory of Groups*, volume 148. Springer Science & Business Media, 2012.

Lones Smith and Peter Sorensen. Pathological outcomes of observational learning, 1996. Available at `http://hdl.handle.net/1721.1/64049`.

Lones Smith and Peter Sorensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

Peter Norman Sorensen. *Rational social learning*. PhD thesis, Massachusetts Institute of Technology, 1996.

John Stillwell. *Classical topology and combinatorial group theory*, volume 72. Springer Science & Business Media, 2012.

M.L. Sylow. Théorèmes sur les groupes de substitutions. *Mathematische Annalen*, 5:584–594, 1872.

Hal R. Varian. Equity, envy, and efficiency. *Journal of Economic Theory*, 9:63–91, 1974.

Xavier Vives. How fast do rational agents learn? *The Review of Economic Studies*, 60(2):329–347, 1993.

Xavier Vives. *Information and learning in markets: the impact of market microstructure*. Princeton University Press, 2010.

Helmut Wielandt. *Finite permutation groups*. Academic Press, 2014.

Jianbo Zhang. Strategic delay and the onset of investment cascades. *The RAND Journal of Economics*, pages 188–205, 1997.

Karol Zyczkowski and Wojciech Slomczynski. Square root voting system, optimal threshold and π. In *Voting Power and Procedures*, pages 127–146. Springer, 2014.