Decision Making in Multi-Agent Groups

by
Megan Stickler

A dissertation submitted to the Department of Mathematics,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Mathematics

Chair of Committee: Krešimir Josić

Co-Chair of Committee: William Ott

Committee Member: Zachary Kilpatrik

Committee Member: Andrew Torok

University of Houston
August 2022

ACKNOWLEDGMENTS

To my advisors, committee, and project members - profoundly, thank you. You've been amazing. Every time someone asks, I tell them how lucky I feel to have gotten to work with you. To my family - thank you also. You know this wouldn't have happened without you! To David - thank you for always being up for talking things out and helping me get untangled. You have the best ideas. To Jen - thank you for being there through everything. Your support has meant a lot. Aldrin - thank you too for your enthusiastic contributions, Buzzy Boy. Your spelling will improve.

ABSTRACT

We explore the ways temporally structured private and social information shape collective decisions. In our first model, we consider a network of rational agents who independently accumulate private evidence that triggers a decision upon reaching a threshold. When seen by the whole network, the first agent's choice initiates a wave of new decisions but later decisions have less impact. In homogeneous networks, the overall probability of a randomly selected agent in such groups making a correct decision is bounded from above because of the impact of the first decider's choice.

In heterogeneous networks, the first decisions are made quickly by impulsive individuals who needed little evidence to make a choice. However, these early decisions, even when wrong, reveal the correct options to nearly everyone else. We conclude that groups comprised of diverse individuals can make more efficient decisions than homogeneous ones.

However, when making decisions, we often rely on a mix of information that we have acquired individually and information that is commonly available. In our second model, we neglect the effect of social information exchange to consider whether the simple fact of information having an individual or a common source affects the quality of decisions. Multiple non-interacting agents make observations, some common and some private, and decide between two options when they have gathered sufficient information to reach one of two symmetric thresholds. In the presence of a mix of common and individual observations, the first agent to reach threshold is less likely to make the correct choice than the first agent reaching threshold when all observations are private or all observations are common. We explain this counterintuitive observation, and conclude that access to common information decreases accuracy for those whose early private information coincides with the common information.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.1 Background

Every day people make decisions, both monumental and trivial, based on their observations and internal biases. Predicting decisions of individuals and large groups of people is a chief concern in many area of applied science including sociology, economics, political science, and ecology. While people's deliberation and commitment processes may be extremely complex and variable, mathematical models have found success in predicting general trends of people performing simple decisions [46]. Modern approaches to quantitative decision-making theory must, at a minimum, involve a decision rule that acts on something concrete. Decisions may be based simply on an observer's bias or on a single observation. For example, one of the oldest political science theorems, the Marquis de Condorcet's Jury Theorem, gives a model of voting behavior in which each voter has a certain probability of voting for the correct choice without reference to processing any evidence or observations. First published in 1785, it formally proves that, under a majority rule, adding more voters will, in the limit of infinitely many voters, always result in a vote in accordance with the environmental state [28].

More modern and complicated voting models still regularly refer to the Condorcet Jury Theorem to provide a benchmark for performance [61, 9]. However, these quantitative models tend to find conditions under which the Jury Theorem fails to hold, often by including complexities the Theorem failed to consider such as ongoing interactions between deciders and environmental sources of information or between the deciders themselves. Thus, while much of the strength of the Condorcet Jury Theorem lies in its simplicity, the extent to which its premises reflect reality is an open question.

Some quantitative decision making models describe deciders that make multiple observations before choosing. Often, this type of model is constructed by layering a decision making rule on top of some evidence accumulation model. For instance, a decision may be triggered when the accumulated evidence reaches some predetermined threshold (a *free response* model), or when evidence has been gathered for some externally determined length of time (an *interrogation* model) [37, 89, 11, 46]. If observations are made of other agents' beliefs or decisions prior to making a decision, a model is said to be an *opinion sharing* or *opinion exchange* model.

It is important to note that a distinction exists between evidence accumulation, opinion sharing, and decision making models as well as between models of individual agents and agents in a group. Evidence accumulation models describe the quantitative processes whereby agents acquire and accumulate multiple pieces of information, either in discrete segments or a continual stream. Opinion exchange models describe processes by which agents gain some knowledge or impression of each other's beliefs and use them to update their own beliefs. Decision-making models involve applying some rule to one's own belief state to acquire a decision.

Any particular model may include any number of these features. For instance, the DeGroot social learning model is an opinion sharing model that does not involve evidence accumulation or decision making. Rather, its agents simply conduct diffusive belief sharing which eventually reaches a steady state [29]. The 2021 paper by Denter *et al.* studying media bias and correlation neglect describes a model that involves evidence accumulation, opinion sharing, and decision making [32]. Ratcliff's 1978 version of the drift-diffusion model for memory retrieval is an evidence accumulation and decision making model, but not an opinion exchange model [84]. Condorcet's Jury Theorem describes a purely decision making model [28]. Granovetter's threshold model, which describes the actions of proportions of populations rather than of individual agents, may also be considered a purely decision making model [49]. The model described and studied by Caginalp and Doiron (2017) involves stochastic evidence accumulation and then the sharing of observers' decisions with one another once these decisions have been made [17].

1.2 Evidence accumulation

Evidence accumulating decision making models have often been developed in the context of the two-alternative forced choice task (TAFC). First detailed by Gustav Fechner in 1889, the TAFC provides a convenient simplifying framework that has been widely applied and studied in the psychological and neurological literature [11, 46, 85, 86, 39]. In TAFC tasks, subjects are required to choose between two options after making noisy observations. With appropriate design choices, the TAFC task framework can be used to study a variety of phenomena such as preference, recognition, and discrimination [4]. While multiple mathematical models have been developed to describe choices made in a TAFC task [16, 31, 105], the drift-diffusion model (DDM) of evidence accumulation and decision making has much to recommend it. In the DDM, observers gather evidence according to a stochastic differential equation whose deterministic part represents the direction of evidence and potentially some evidence discounting and whose white noise term represents variability in observations.

The DDM arises as the continuous analog of a discrete random walk. The use of random walks in decision making gained traction during World War II with the development of the sequential probability ratio test (SPRT). The SPRT constructs a random walk whose moves are determined by the likelihood of successive observations and is used to choose between two different hypotheses. This test was developed independently by George Barnard in England, Abraham Wald in America, and Alan Turing (in documents that were declassified decades later) [30, 108, 48]. Both the SPRT and DDM have been shown to produce an optimal relationship between the speed and accuracy of decisions. Variants of the DDM retain this optimality under an appropriate choice of parameters [109, 11, 98].

In the following decades, both the DDM and the SPRT were shown to capture well the accuracy and reaction times in humans performing TAFC tasks in laboratory conditions [84, 87, 67, 99, 106]. For example, first passage time distributions for the DDM capture the long tails in human reaction time distributions [84, 94, 4, 86].

The DDM has also been shown to be useful for distinguishing behavioral components such as increased caution from physical components such as slow motor response in experimental performance [75, 47]. Research suggests that the parameters of the DDM do map well to various aspects of TAFC tasks: threshold to accuracy motivation, drift rates to discrimination difficulty, and a biased prior toward rewarded choices, among others [107, 64]. Modifications to the standard DDM allow it to be applied quite broadly. (See Figure 1.1 adapted from [87] for a sketch of the taxonomy of sequential sampling and drift-diffusion evidence accumulation strategies.)

More recent research uses the DDM to model neuron firing patterns in visual discrimination tasks such as the random dot motion discrimination task[1] for humans and visual word recognition tasks for humans and for non-human primates [93, 92, 40, 46, 45, 95]. Characteristics of the

[1]The random dot motion discrimination (RDMD) task involves a subject observing a movie of randomly moving dots for a short period of time and determining the mean direction the dots are drifting. The difficult of the task can be controlled by how coherently the dots drift.

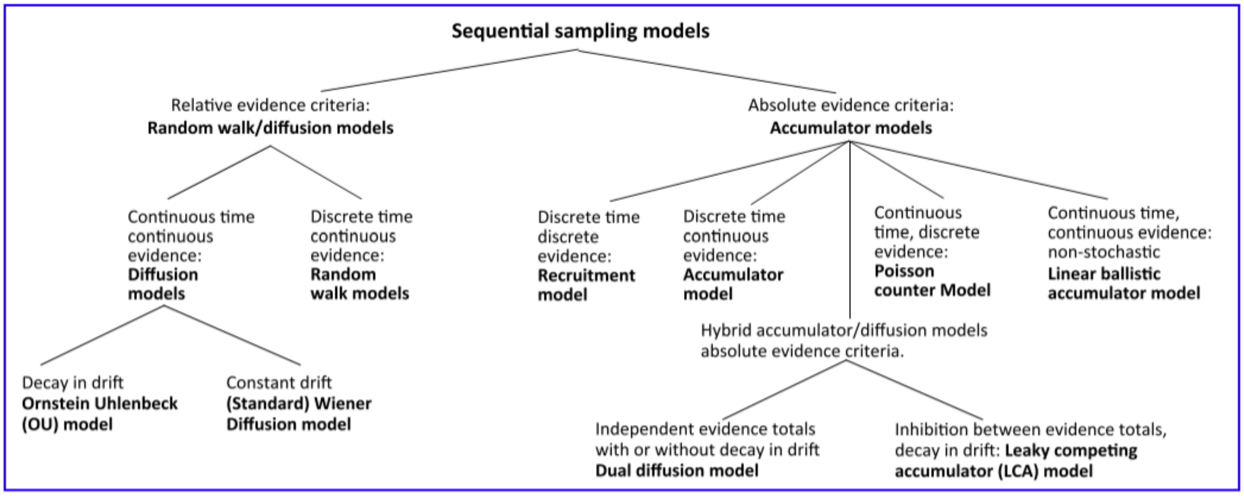

Figure 1.1: A family tree of evidence accumulation methods deriving from the sequential probability ratio test (SPRT). Adapted from Ratcliff *et al.* (2016) [87].

model such as the speed-accuracy tradeoff (SAT) between the slow, likely accurate decisions and fast, risky decisions may have a physiological basis [12]. There is some evidence that different aspects of the model may be associated with different brain structures; for example, evidence accumulation with a fronto-parietal network or an inferior-temporal network, or decision threshold with a fronto-basal ganglia network [77, 103, 55].

1.3 Social models

Most of the studies listed above concern decision making involving only one evidence-accumulating, decision-making agent. However, social decisions are of great interest in fields such as economics, politics, biology, and psychology [24, 37, 65, 113]. Relevant questions about the manner in which groups affect the speed, accuracy, and coherence of decisions have complicated answers and results vary by model and assumptions [27, 110, 63]. In Condorcet-type groups, decision makers reach decisions individually and a rule is applied to the collection of individual decisions to produce a collective decision. In network-type groups, individuals interact with each other to produce individual and collective decisions [98]

In biology, social decision making may describe collective decisions such as habitat relocation or finding of food sources by ants, honeybees, schools of fish, troops of baboons, and others. In each of these examples, individuals possess limited information but pool it to generate coherent and adaptive group responses [91, 71, 22, 100]. These consensus decisions have been studied using game-theoretic models [23, 66], leadership models in which certain individuals have outsized influence [26, 35, 82], quorum models in which a majority rule holds [111, 102], and even Ising-type models that represent individuals with spin states that may propagate through a population [52]. While many of these models exhibit a speed-cohesion tradeoff in which achieving consensus takes longer for larger groups [42, 96, 22], group decisions may remain highly efficient even in situations where bias or beliefs are in conflict [34, 25].

Biological models often assume a *Bayesian* (rational) decision making approach. Interestingly, experimental results suggest that common mechanisms underlie both individual decisions such as food choice, and social decisions such as resource allocation and that animals use both individual and social evidence when making decisions [65, 69]. Even in animals like fish whose movement rules can be described simply, it can be shown that the decisions made by the collective follow Bayesian principles [79, 3]. However, across-trial variability in performance may be the result of varying ability or inclination of subjects to follow Bayesian principles [8]. The imprecise nature of information aggregation is potentially another major contributor to sub-optimal decisions [33, 18].

Many theoretic models also depend strongly on the ability of agents (animal or human) to clearly distinguish numerosity. Fortunately, observations of both primates and humans reveal some type of adequate number representation [81]. Many types of animals have some ability to discriminate between larger and smaller groups of their conspecifics, even when the ratio of group sizes is relatively close [10]. Studies have found that humans and non-human primates tend to follow a version of Weber's law: their ability to discriminate numerosity depends on the ratio of the sizes of the groups rather than their absolute sizes [36]. Human adults appear to be able to discriminate groups with a ratio of ratios of 1.15 and find crossmodal comparisons in which they compare groups perceived using one sense with groups perceived using another sense only slightly more difficult than intramodal comparisons which use the same sense for both [7].

Humans and non-human primates also appear to possess some intuitive statistical reasoning potentially associated with a biological analogue magnitude system. Both chimpanzees and adult humans can infer the makeup of a sample from the makeup of a population: When selecting between samples from two transparent containers, both species choose the sample that came from the container with a higher ratio of favorable to unfavorable contents [36].

Models from the economics literature often focus on the behavior of individuals or populations in the context of markets. Granovetter's threshold model and Banerjee's herding model have been used to study adoption of behaviors and products by populations [49, 6]. Knowledge may be transmitted in markets by decisions to purchase or produce or by price adjustments. [70]. Market models may demonstrate such behaviors as cascades and opinion convergence [112, 43]. Many economic models consider the effect of network structure on the propagation of decisions through a system [43, 112, 1, 76, 74].

In psychology, models tend to focus on the nature of interactions between parties, such as who listens to whom and why, while comparing actual experimental results with theoretic optimal results [5, 63]. In addition to the TAFC framework, decision field theory is an oft-used dynamic-cognitive approach to decision making and preferential choice that presents the evolution of preferences as a dynamical system [16, 15].

According to experimental investigation of social learning, the rate at which individuals copy each other increases with number of individuals who *demonstrate* or model a behavior, demonstrator consensus (which is enhanced in larger groups), uncertainty on the part of the copier, and the cost of individual learning. Individuals in the minority are more likely to change their behavior than individuals in the majority [72]. Studies differentiate between conformist transmission, in which majority behaviors are copied at a rate higher than their appearance in the population, and unbiased transmission, in which majority behaviors are copied at a rate commensurate with their rate of appearance in the population [13, 53].

Interestingly, some experimental results suggest different regions of the brain are involved in processing socially and individually acquired information [78]. Group decisions may be more reliable than individual decisions for groups of risk-accepting individuals [51]. While [54] gives theoretical results finding that Bayesian calculations for groups are NP hard, it may nevertheless be that humans can make essentially Bayesian decisions regarding their expectation of group decisions [62].

Several authors have taken the natural step of extending the single-agent TAFC to encompass group decisions [63] and some studies consider interactions between agents [98, 17, 60]. There is an evolutionary approach to social learning that incorporates both individual and social information, but not time-based evidence accumulation [80].

In this dissertation, we consider agents who accumulate evidence over time both from individual sources and by their observations of other agents' decisions. Initially we explore the effects of Bayesian observations of other agents' decisions on agents who are also collecting their own information according to a DDM (the 'Standard Diffusion Model' in Figure 1.1) whose stochastic term is independent of the stochastic terms of all other agents. Each agent applies this information to their own TAFC to produce decisions.

Holding the drift term constant, we examine the differing impact of Bayesian and non-Bayesian social updating by considering groups in which all members have the same threshold, groups in which members have differing thresholds but update in a manner as though all group members had the same threshold, and groups in which members have differing thresholds and update in an omniscient manner that reflects knowledge of these differing thresholds. We find that the probability of some random group agent reaching a decision consistent with environmental evidence (a *correct* decision) is optimized when agents perform omniscient social updating in groups whose members have differing thresholds.

In our later work, we explore what happens when individual agents in groups collect their own information according to the SPRT (the 'Random Walk Model' in Figure 1.1) and this information is partially correlated with the information being collected by other agents. In particular, we examine how the accuracy of agents' decisions with fixed belief thresholds depends on the order in which they make decisions.

1.4 Correlated information models

The study of the effects of correlated information sources on the efficacy of collective decisions or the decisions of group members deciding individually has a long history in the economics and political science literature. For example, Banerjee's herding model, which studies the effects of social information, can also be interpreted as a study of the effects of correlated observations: when attempting to infer the private evidence of those who chose before them, later deciders must consider that middle deciders were themselves observing the same set of early deciders [6]. In such a setting, knowledge of the correlated data comprised of the decisions of early deciders may both completely mask the private observations of the middle deciders by making these private observations irrelevant to middle deciders' actions and prevent the actual decisions made by the middle deciders from providing any new information. More complicated voting models such as that discussed in [32] may include both accumulation of correlated evidence and opinion exchange based on this correlated evidence before a vote is taken.

In models and experiments, correlation often appears in one of two ways. In the first, an agent receives two or more pieces of information which have some degree of correlation with each other. In the second, an agent receives a single piece of evidence from another agent (either as an opinion or an observation of another agent's decision), but that shared evidence was formed from multiple correlated individual observations. Models studying the first type of correlation may use single agents or multiple agents; models studying the second type must necessarily involve multiple agents. For example, Spiegler's 2016 study presents a model of a single agent who is deciding whether or not to pursue a diet based on observations of an irrelevant variable that has a random degree of correlation with several actually relevant variables [97]; this is an example of a single agent model using correlation of the first type.

Qualitatively, early political scientists often expected that group decisions would be more deliberative than individual decisions, taking advantage of the wisdom of crowds to balance out extremist viewpoints; indeed, the proponents of a 'free marketplace of ideas' used this classical ideal of democracy as a justification [56]. However, quantitative models have suggested that deliberation (modeled by opinion exchange) can instead lead to viewpoints becoming more extreme and experimental evidence has consistently shown that real human political systems very rarely emulate the classical democratic ideal [14, 90, 104]. According to quantitative models, this extremism as a result of opinion exchange is quite rational [44].

Another strain of correlation studies in the political science realm examines the effect of correlated sources of information on the outcomes of voting models, using the outcomes expected by the Condorcet Jury Theorem as a benchmark. Some find that correlated information from single sources such as ‘experts’ or news outlets can result in groups that reach an optimal decision with probability less than one, failing the promise of the Jury Theorem. This failure may be theoretically avoidable but robust to circumstance in laboratory settings [61], or appear only under certain conditions such as a suboptimal prior and a high degree of network connectivity [32].

In many studies *correlation neglect*, which occurs when agents fail to consider that both they and other agents have access to the same information when updating their beliefs based on actions or opinions of other agents, has been shown to have interesting effects. For example, while increased extremism is a likely result of Bayesian social updating, non-Bayesian social updating that neglects correlation has been shown to increase the probability of individual recovery from a non-optimal prior in populations with correlated evidence sources [68, 73].

Experimentally, correlation neglect is quite plausible. It may come in the form of ‘double-counting’, wherein people count the re-telling of a news story they have already heard as a separate incidence of stories of that nature [38]. While humans have the intellectual capacity to dismiss the retelling, studies suggest that most are unlikely to put forth the effort required to distinguish the effect of correlation in the information they receive without some incentive or reminder commensurate with the difficulty of the task [38, 19, 20].

For particularly difficult tasks, people are unlikely to attempt to disentangle and dismiss the effects of correlation in their information, even when the potential reward is large [58]. Experiments show that a majority of naïve subjects do not recognize even particularly prominent results of correlation such as information cascades [50]. However, trained subjects, such as professionals from the Chicago Board of Trade, are both able and likely to compensate for cascade effects in the correlated evidence they gather from the decisions of other professionals [2].

Our work does not fit exactly into the above framework: the type of correlation we study is neither type one, in which a single agent's observations are correlated with each other, nor type two, in which an agent receives information from a neighbor whose information came from correlated sources. Rather, we examine an evidence accumulation model based on the SPRT in which some observations are made in common with all other agents, but each agent is unaware of the presence of other agents. We examine the effect of this common evidence on the accuracy of the agent's decisions.

Specifically, we ask "Does the mere presence of a common information source (such as a news station) in the environment have an effect on decision quality given fixed individual belief decision thresholds, even in the absence of social influence?" We find that common information, even when of the same quality as individual information, tends to dominate the direction of the first decisions made by group members, especially in large groups. This results in decisions whose accuracy is predicted by the amount of information available in the common observations rather than the total amount of information available from both common and individual sources.

Chapter 2

Combining evidence accumulation and social decision making: A model

Evidence accumulation refers to the process by which one makes a sequence of observations over time and interprets those observations to inform their choices. In our mathematical models, agents can noisily observe the state of the environment (*e.g.*, swirling leaves inform one of the wind direction, people entering a cafe with raincoats/umbrellas suggests it's raining outside), or other agents (*e.g.*, if people are buying a stock it may be because the value is about to increase).

When observing other agents, observations can provide information about other agents' private evidence (as in opinion exchange models) or other agents' choices. We can assume that agents have a probabilistic model of the world, and interpret observations. If agents use this model to compute the probabilities that the world is in a particular state, we say that they are *Bayesian* or rational. Agents can also use heuristic rules to make inferences about the state of the world, and we call such agents *non-Bayesian*. An evidence accumulation model can be used as a basis for a decision making model if, for instance, agents make decisions based on some rule that takes into account accumulated evidence or other attributes such as total time spent on the observations or total amount of evidence available in the environment.

Our work focuses on decision making models in which a decision rule is applied to agents' accumulated evidence. The evidence accumulated is based on environmental observations and, depending on the precise model under consideration, also observations about other agents' decision states. We look both at models where observations are made at discrete timesteps ('*discrete models*') and at models where observations are made continuously over time ('*continuous models*').

For our purposes, the goal of evidence gathering is to determine which of two hypothetical environmental states H^+, H^- is more likely given the available evidence. Evidence in favor of one environmental state disfavors the other. We will also assume that agents make decisions when the accumulated evidence exceeds predetermined thresholds,which provide boundaries on a symmetric random walk.

2.1 Discrete accumulation

Our discrete model of evidence accumulation derives from the sequential probability ratio test (SPRT) introduced in Wald's 1945 paper as a way to make decisions based on evidence accumulated from multiple observations over time [108]. This test is quite popular; in 1948 Wald and Wolfowitz showed it to be optimal in that, given a desired degree of accuracy, the SPRT achieves this accuracy with a minimum average number of samples [109].

In Wald's original description of the SPRT, one supposes observations are drawn from one of two distinct probability distributions $p_0(x)$ and $p_1(x)$. Given a series of independent observations, one attempts to determine whether they have been drawn from the distribution $p_0(x)$ (satisfying hypothesis H_0, which supposes $p_0(x)$ is the environmental distribution) or the distribution $p_1(x)$ (satisfying hypothesis H_1). We refer to H^0 and H^1 as *environmental states*: the environmental state is H^0 if hypothesis H_0 is satisfied, or H^1 if hypothesis H_1 is satisfied.

In our work, we assume that p_0 and p_1 are symmetric across $x = 0$ so that $p_0(x) = p_1(-x)$. We assume further that for one of the distributions the bulk of the mass occurs for $x > 0$, and denote this distribution $p_+(x)$, which produces observations when hypothesis H_+ is true and the environmental state is H^+. Because of symmetry, the other distribution has most of its mass at $x < 0$, and we denote this distribution $p_-(x)$, which produces observations when the hypothesis H_- is true and the environmental state is H^-. Following these assumptions, the conditional probabilities that an agent makes observation x are $P(x \mid H^+) = p_+(x)$ and $P(x \mid H^-) = p_-(x)$. Throughout this dissertation, we will abbreviate the conditioning on H^+ and H^- by defining

$$P^+(\cdot) := P(\cdot \mid H^+)$$

and

$$P^-(\cdot) := P(\cdot \mid H^-).$$

In Wald and Wolfowitz (1948) the criterion for decision-making according to the SPRT uses a likelihood ratio and is given as follows [109]:

Take two positive numbers A and B such that $A > 1$ and $B < 1$. We will call these numbers A and B *thresholds*. Let the *trajectory* ξ be an infinite sequence of observations drawn from $p_\pm$ (if $H^\pm$ is true):

$$\xi = \{x_1, x_2, \dots . x_t \dots\}$$

and let $\xi_{1:t}$, ξ_i be subsets of that sequence:

$$\xi_{1:t} = \{x_1, x_2, \dots . x_t\}; \text{ and } \xi_i = \{x_i\}.$$

Then we have by conditional independence of samples:

$$P^{\pm}(\xi_{1:t}) = \prod_{i=1}^{t} P^{\pm}(\xi_i) = \prod_{i=1}^{t} p_{\pm}(x_i),$$

We define the function $T(\xi)$ on the set of all possible sequences of observations ξ, so that $T = \hat{T}$ if $\hat{T}$ is the minimal value of t such that either

$$\frac{P^{+}(\xi_{1:t})}{P^{-}(\xi_{1:t})} \geq A \text{ or } \leq B.$$

If

$$\frac{P^{+}(\xi_{1:t})}{P^{-}(\xi_{1:t})} \geq A$$

then the decision-making agent decides the environmental state is H^{+}; else if

$$\frac{P^{+}(\xi_{1:t})}{P^{-}(\xi_{1:t})} \leq B$$

then hypothesis H^{-} is accepted. In either case, given trajectory ξ, the decision rule is triggered at *decision time* $T(\xi)$. According to [109], the distributions p_0, p_1 (or in our notation, p_-, p_+) ought to be chosen so that T is finite with probability 1.

In our work we use the more convenient log-likelihood ratio criterion to determine our decision time T.

Given a single observation x, the log-likelihood ratio is given by

$$\text{LLR}(x) := \log \frac{P^{+}(x)}{P^{-}(x)}.$$

For a sequence of observations $\xi_{1:t}$, we have

$$\begin{aligned}\mathrm{LLR}(\xi_{1:t}) &:= \log \frac{P^+(\xi_{1:t})}{P^-(\xi_{1:t})} = \log \frac{\prod_{i=1}^t P^+(\xi_i)}{\prod_{i=1}^t P^-(\xi_i)} \\ &= \log \prod_{i=1}^t \frac{P^+(\xi_i)}{P^-(\xi_i)} = \sum_{i=1}^t \log \frac{P^+(\xi_i)}{P^-(\xi_i)} \\ &= \sum_{i=1}^t \mathrm{LLR}(\xi_i),\end{aligned}$$

due to the conditional dependence of observations ξ_j and laws of logarithms.

We also adjust our thresholds

$$A^* = \log A;$$

$$B^* = \log B$$

so that we have $A^* > 0$ and $B^* < 0$. Accordingly, the adjusted decision rule is that a decision is made at time T, where T is the minimum time t such that

$$\mathrm{LLR}(\xi_{1:t}) \geq A^* \text{ or } \mathrm{LLR}(\xi_{1:t}) \leq B^*.$$

As before, we have that when

$$\mathrm{LLR}(\xi_{1:t}) \geq A^*$$

the decision-making agent decides the environmental state is H^+, and if

$$\mathrm{LLR}(\xi_{1:t}) \leq B^*,$$

the decision-making agent decides the environmental state is H^-. As in the case of the likelihood ratio criterion, we still have that T is finite with probability 1.

In our models, we assume that the thresholds A^* and B^* are symmetric about 0 and denote them by θ, $-\theta$ respectively. We consider ideal agents who know the probability of drawing each observation ξ_i in either environmental state (we assume agents know $P^{\pm}(\xi_i)$ exactly). These ideal agents do not forget any observations that they have made (the evidence accumulation is not *leaky*). Furthermore, we assume that the decision-making agent believes each of the two environmental states to be equally likely *a priori*, so that $P(H^+) = P(H^-) = 1/2$.

Under these assumptions, we may say that an agent's accumulated evidence y at time t is given by the LLR of its observations up to time t:

$$y(t) = \text{LLR}(\xi_{1:t}). \tag{2.1}$$

We refer to this accumulated evidence $y(t)$ as the *belief* of the agent at time t. The assumption of a flat prior implies that $y(0) = 0$.

2.2 Continuous accumulation

Taking the limit of infinitely rapid and infinitesimally weak observations in the discrete evidence accumulation and decision-making model described in the last section gives rise to an analogous continuous model. Given the LLR dynamics described above, we have that between timestep $t-1$ and timestep t, the change in the belief of our agent may be given as

$$\Delta y = \text{LLR}(\xi_t) = \log \frac{P^+(\xi_t)}{P^-(\xi_t)} = \log \frac{p_+(x_t)}{p_-(x_t)}.$$

We introduce an expectation for the value of ξ_t based on the environmental state H:

$$\Delta y = \mathbb{E}_{\xi_t}\left[\log \frac{p_+(x_t)}{p_-(x_t)}|H\right] + \log \frac{p_+(x_t)}{p_-(x_t)} - \mathbb{E}_{\xi_t}\left[\log \frac{p_+(x_t)}{p_-(x_t)}|H\right].$$

If we then scale the expectation and variance by the size of the timestep Δt and also take $\Delta t \to 0$, we obtain

$$dy = h(t)dt + \rho(t)dW,$$

where W is a standard Wiener process and $h(t)$, $\rho^2(t)$ are defined as

$$h(t) = \frac{1}{\Delta t}\mathbb{E}_{\xi_t}\left[\log\frac{p_+(x_t)}{p_-(x_t)}|H\right];$$
$$\rho^2(t) = \frac{1}{\Delta t}\,\mathrm{Var}_{\xi_t}\left[\log\frac{p_+(x_t)}{p_-(x_t)}|H\right].$$

Using $p_\pm(x) = \frac{1}{\sqrt{2\pi\Delta t\sigma^2}}e^{-(x-\Delta t\mu_\pm)^2/(2\Delta t\sigma^2)}$, we have

$$h(t) = \pm\frac{(\mu_+ - \mu_-)^2}{2\sigma^2}$$
$$\rho^2(t) = \frac{(\mu_+ - \mu_-)^2}{\sigma^2}.$$

Recalling that we chose $p_\pm(x)$ to mirror each other across $x = 0$, we can let $\mu_+ = 1$, $\mu_- = -1$, and $\sigma^2 = 1$.

This then gives us a continuous model where evidence is accrued with the stochastic drift-diffusion equation

$$dy = \alpha dt + \sqrt{2}dW,$$

where W is a standard Wiener process and α depends on the true environmental state:

$$\alpha = \begin{cases} 1 & H^+ \\ -1 & H^- \end{cases}.$$

We retain the flat prior so that evidence accumulation again begins with $y(0) = 0$ and accretes with

$y(t) = \int_0^t \frac{dy}{dt}$.

Again, a decision is made at time T, where T is the minimum time t such that $y(t) = \theta$ or $y(t) = -\theta$. Reaching the positive threshold θ results in the agent deciding the environmental state is H^+; reaching the negative threshold $-\theta$ results in the agent deciding the environmental state is H^-.

Without loss of generality, we will henceforward assume that the true environmental state is H^+. This state is not known to the agent or agents. Thus, we will often refer to an agent's decision that the environmental state is H^+ as a *correct* or *accurate* decision and to an agent's deciding that the environmental state is H^- as a *wrong* decision. Thus, in our models $\alpha = 1$ and

$$dy = dt + \sqrt{2}dW. \tag{2.2}$$

2.3 Multiple agents

We recall that [109] showed the SPRT to be optimal in that it required a minimum number of samples to achieve a desired degree of accuracy; that is, given the restriction of a particular degree of accuracy, an SPRT model will (on average) result in the fastest possible decision. It can be shown that, given symmetric thresholds $-\theta < 0 < \theta$, the expected amount of time before a single agent who is accumulating evidence according to the continuous Eq. (2.2) reaches any threshold is

$$\mathbb{E}[T] = \theta \tanh(\theta/2). \tag{2.3}$$

The probability that the decision that agent made is accurate is

$$P^+\Big(y(T) = \theta\Big) = \frac{1}{1+e^{-\theta}}. \tag{2.4}$$

This is the probability (given a true environmental state of H^+) that the first threshold the agent reaches will be the positive one.

Thus, we see that both the speed and the accuracy of the agent's decision depend on the size of the threshold, θ: larger θ values will , on average, result in slower, more accurate decisions. Smaller θ values, on average, give faster, less accurate decisions. This dichotomy is known as the speed-accuracy tradeoff (SAT). We explore how moving from a single agent model to a multiple agent model can affect the SAT.

Our work focuses on two separate decision making models which build on the process discussed thus far. The first, which focuses on information-sharing within a large clique and which we explore more in Chapters 3 and 4, is detailed below. The other, which focuses on the role of correlated observations in shaping the accuracy of individuals in a group, is detailed in Chapter 5 and explored further in Chapter 6.

2.4 Continuous evidence accumulation model with social evidence

The first model we consider is a continuous evidence-accumulation model analogous to the discrete evidence accumulation model described in Section 8 of [60]. We assume that decisions are made by a group of N agents. Each of these agents has two sources of information, private and social, which are combined to form that agent's belief about the state of the world at any given time.

Private information comes from observing the environment directly and is accumulated by each agent according to the SDE given in Eq. (2.2). Social information comes from observing the decision states of other agents in the group. Note that this is not an opinion-exchange model: private evidence is never exchanged directly. Rather, each agent makes inferences about the amount of other agents' private evidence based on observations of other agents' *decision states*.

An agent's decision state is a parameter that indicates whether that agent has yet reached a decision at a given time and if so, what that decision was. This is equivalent to saying the decision state indicates whether that agent's total belief has reached either threshold and if so, which threshold it reached. We assume that agents stop accumulating evidence when they reach a threshold, so decision states may change only once: once made, a decision is immutable.

We describe the decision state notationally thus:

In a group of N agents, suppose agent i ($1 \leq i \leq N$) has a total belief at time t given by $y_i(t)$ and a decision threshold θ_i. Then agent i's decision state may be given with

$$d\Big(y_i(t)\Big) = \begin{cases} 0 & \big|y_i(t)\big| < \theta_i \\ 1 & y_i(t) \geq \theta_i \\ -1 & y_i(t) \leq -\theta_i \end{cases} . \tag{2.5}$$

Let T_i be the earliest time t such that $\big|y_i(T_i)\big| \geq \theta_i$, the time when agent i's total belief first reaches a threshold. This corresponds to the time that agent i makes a decision. Since decisions are immutable (we choose for our models that evidence accumulation ceases once a decision has been made), we have that for $t < T_i$, $d(y_i(t)) = 0$ while for $t \geq T_i$, $\big|d(y_i(t))\big| = 1$. Later, we will heavily use $T = \min_{1 \leq i \leq N} (T_i)$, the *first decision time*, *i.e.* the time of the first decision in the group.

Following the model given in Section 8 of [60], the network arrangement of our group in this model is all-to-all. That is, each agent in the group has a perfect view of the decision state of each other agent. Moreover, all agents are completely rational and account exactly for all social information they know other agents have received when forming their inferences about other agents' private evidence amounts. Each agent is also aware that all other agents accumulate private evidence according to Eq. (2.2).

Our model, then, consists of a group of N agents constantly observing both their environment and each other's decision states and making inferences about the true environmental state based on both of these sources of information. These inferences yield evidence which each agent accumulates to form their total belief at any given time.

We assume that any time new social evidence becomes available, the time pauses until all social evidence is incorporated. We 'pause time' by halting the integration of private evidence until all social evidence is exchanged between the agents. During social evidence exchange, the evidence provided by first decision will trigger some subgroup of the undecided agents to make a decision. The evidence provided by the decisions of those in this subgroup (*wave*) and non-decisions of those not in the wave will then trigger decisions on the part of successive subgroups (waves) until either all agents have reached a decision or the evidence provided by the last wave is insufficient to trigger any more decisions. This process is explained in detail in Section 2.5.

The total belief of agent i at time t is the sum of their private and social information:

$$y_i(t) = y_{priv}^{(i)}(t) + y_{soc}^{(i)}(t). \tag{2.6}$$

Agent i continues to accumulate evidence until their total belief crosses one of two symmetric thresholds $-\theta_i, \theta_i$.

Since all agents can see each other, at each point in time $y_{soc}^{(i)}(t)$, the amount of social information available to agent i, is known by all other agents. Using this and agent i's decision state, other agents may infer that agent i's private evidence $y_{priv}^{(i)}(t)$ is in some interval (a,b). We defer discussion of the method of this inference to Section 2.5. The amount of evidence an agent $j \neq i$

acquires from observing or knowing that $y^{(i)}_{priv}(t) \in (a,b)$ is given by

$$\begin{aligned} Soc_i(t) &= \mathrm{LLR}\left(y^{(i)}_{priv}(t) \in (a,b) \right) \\ &= \log \frac{P^{+}\left(y^{(i)}_{priv}(t) \in (a,b) \right)}{P^{-}\left(y^{(i)}_{priv}(t) \in (a,b) \right)} \end{aligned} \tag{2.7}$$

The total amount of social evidence in the group is

$$Soc(t) = \sum_{1 \le j \le N} Soc_j(t),$$

and the amount of social information available to agent i is

$$y^{(i)}_{soc}(t) = Soc(t) - Soc_i(t) = \sum_{j=1, j \ne i}^{N} Soc_j(t)$$

because agent i does not acquire social information from itself.

To calculate $Soc_i(t)$, we recall that each agent accumulates private information according to Eq. (2.2):

$$y^{(i)}_{priv}(t) = \int_0^t (dt + \sqrt{2} dW). \tag{2.8}$$

The private evidence, $y^{(i)}_{priv}(t)$, for $t \le T$, $1 \le i \le N$ is distributed according to the pdf $p^{*}_{\pm}(x,t)$ which gives the distribution of agent's beliefs evolving according to Eq. (2.2) conditioned on no agents' private evidence having left the interval $(-\theta, \theta)$ at any time previous to t.

Then, given an interval (a,b) where $-\theta \le a < b \le \theta$, the probability of a given agent's private evidence being in the interval (a,b) at time $t \le T$ is

$$P^{\pm}\left(y^{(i)}_{priv}(t) \in (a,b) \right) = \int_a^b p^{*}_{\pm}(x,t) dx. \tag{2.9}$$

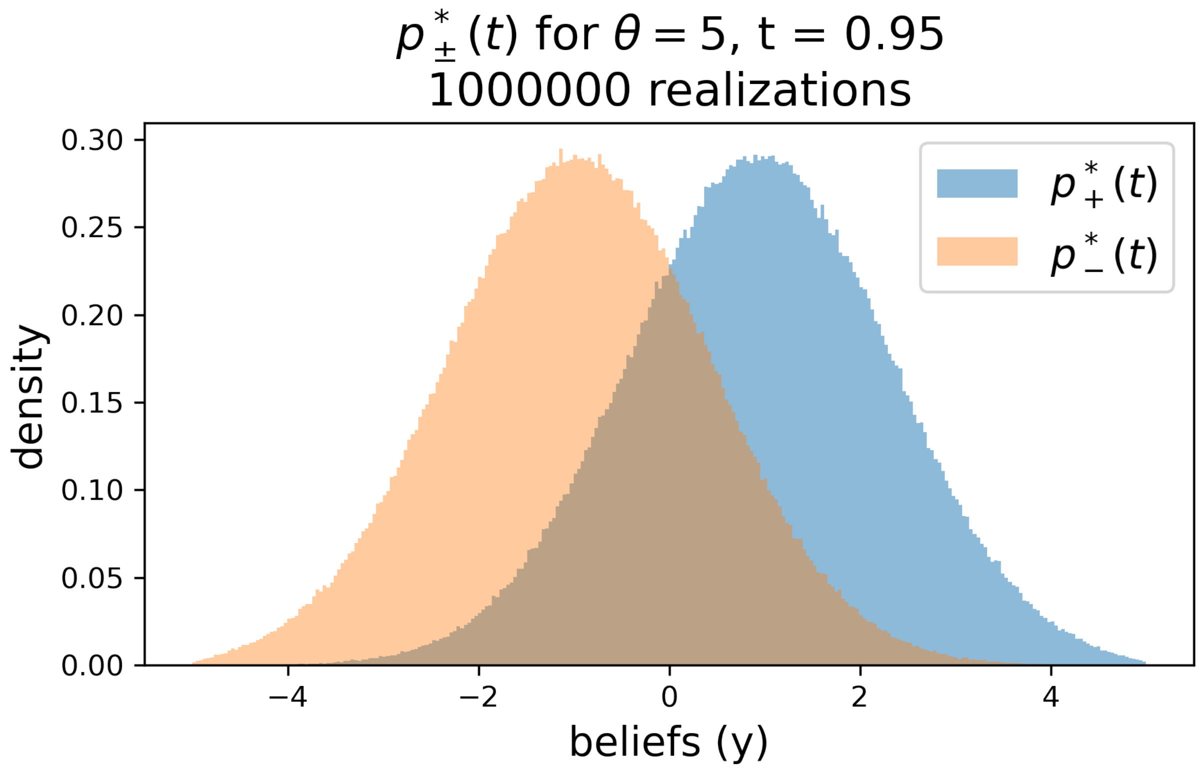

Figure 2.1: Comparison of $p^*_+(t)$ and $p^*_-(t)$. These are the distributions of agents' beliefs at time t conditioned on no agent's belief having left the interval $(-\theta,\theta)$ when beliefs are accumulated according to Eq. (2.2) ($p^*_+(t)$) and when beliefs are accumulated according to an analogous equation with a negative drift term, $y^{(i)}_{priv}(t) = \int_0^t(-dt+\sqrt{2}dW)$ to form the distribution $p^*_-(t)$. The distributions are such that the bulk of $p^*_+(t)$ occurs for $y > 0$, the bulk of $p^*_-(t)$ for $y < 0$, and the distributions mirror each other across 0 ($p^*_+(y,t) = p^*_-(-y,t)$). Figure generated via Monte Carlo methods over 10^6 realizations for values $\theta = 5$, $t = 0.95$.

As we noted earlier, the discrete-case evidence distributions p_+ and p_- are symmetric about 0. If we follow the derivation in the previous sections assuming either H^+ or H^- is the true environmental state, we find that $p^*_\pm(x,t)$ are also symmetric about 0 (see Figure 2.1). Due to this reflective symmetry, we have

$$
\begin{aligned}
P^{+}\left(y_{priv}^{(i)}(t) \in (a,b)\right) &= \int_{a}^{b} p_{+}^{*}(x,t)dx \\
&= \int_{-b}^{-a} p_{-}^{*}(x,t)dx \\
&= P^{-}\left(y_{priv}^{(i)}(t) \in (-b,-a)\right).
\end{aligned}
$$

Using this symmetry, can obtain probabilities P^{-} using only p_{+}^{*}:

$$
\begin{aligned}
Soc_i(t) &= \log \frac{P^{+}\left(y_{priv}^{(i)}(t) \in (a,b)\right)}{P^{-}\left(y_{priv}^{(i)}(t) \in (a,b)\right)} \\
&= \log \frac{P^{+}\left(y_{priv}^{(i)}(t) \in (a,b)\right)}{P^{+}\left(y_{priv}^{(i)}(t) \in (-b,-a)\right)} \\
&= \log \frac{\int_{a}^{b} p_{+}^{*}(x,t)dx}{\int_{-b}^{-a} p_{+}^{*}(x,t)dx}.
\end{aligned}
\tag{2.10}
$$

Next, we explain how the quantity $Soc_i(t)$ evolves as the decision states of agents within the group progress.

2.5 Evolution of social evidence and decision waves

Before the time of the first decision, T, no agent has yet made a decision. Under our model assumption that thresholds are symmetric about 0, non-decisions are uninformative and $Soc_i(t < T) = 0$ for all agents, since it is just as likely that a wrong decision would not be made by a given time as that a right decision would not be made by that same time. Accordingly, prior to the first decision, $Soc_i(t) = Soc(t) = y_{soc}^{(i)}(t) = 0$ for all agents.

Without loss of generality, suppose it is agent 1 who makes the first decision. Then all other agents pause their integration of new private evidence until they have acquired all the social information they can from other agents' changing decision states.

Agents consider social evidence in a sequence of *waves* whose three steps are

1. Decision waves: Any agent who has accumulated sufficient evidence to reach a threshold makes their decision. Any agent deciding at this step will be considered part of the current wave A_W:

$$A_W := \{j \mid |y_j(T^W)| \geq \theta, |y_j(t)| < \theta \text{ for } t < T^W, 1 \leq j \leq N\}.$$

   We refer to the first deciding agent as the 'zeroth wave', A_0. Any agent who decides as a result of social evidence acquired from the first agent's decision will be part of the first wave A_1; those who decide as a result of decisions in the first wave will be part of the second wave A_2, etc.

   We use the time notation T^W to refer to the time at which decisions in wave A_W occur. Using this notation, $T = T^{(0)}$ is the first decision time T at wave 0, the first decision. Similarly, $T^{(1)}$ describes the time at which the first wave decides, $T^{(2)}$ the time at which the second wave decides, and so on.

2. Social information updates: The decision state of all agents provides social information to those agents who have not yet made a decision.

   (a) Update $Soc(T^{W-1})$ to $Soc(T^W)$: The sum of social information from each agent will be $Soc(T^{(0)})$ after the zeroth wave (first decision), $Soc(T^{(1)})$ after the first wave, etc. $Soc_i(T^W)$ is calculated for each agent by finding the interval $(a_{i,W}, b_{i,W})$ in which $y^{(i)}_{priv}(T)$ lies or must have lain before the agent reached a decision.

   (b) Update $y_i(T^{W-1})$ to $y_i(T^W)$: All undecided agents (those who do not yet belong to any wave) update their beliefs with this new social information:

$$y_i(T^{W-1}) = y_{priv}^{(i)}(T) + y_{soc}^{(i)}(T^{W-1})$$

becomes

$$y_i(T^W) = y_{priv}^{(i)}(T) + y_{soc}^{(i)}(T^W).$$

We note that because agents' private information has ceased to accumulate during the rounds of social updating, for all agents j and all waves W we have $y_{priv}^{(j)}(T) = y_{priv}^{(j)}(T^W)$.

When $y_{soc}^{(j)}(T^W)$ is the same for all undecided agents j, we refer to this quantity as the *social increment* following wave W. (See Section 3.6.)

3. If there is some agent i not part of any previous wave whose belief $y_i(T^W)$ is at either threshold, return to step one. Otherwise, social updating is complete and we return to gathering private evidence.

See Figure 2.2 for an illustration of the first and second wave cycles in the case of a wrong first decision. Recall that we have supposed, without loss of generality, that agent 1 makes the first decision, triggering the beginning of the wave cycle. We are now at time $T^{(0)}$ and the zeroth wave contains exactly one agent, agent 1, so that $A_0 = \{1\}$. Next we calculate the social information update $Soc(T^{(0)})$ by finding our interval $(a_{i,0}, b_{i,0})$ for each agent's private evidence and using it in Eq. (2.10).

After this zeroth wave, all agents except agent 1 are still undecided. Because social information

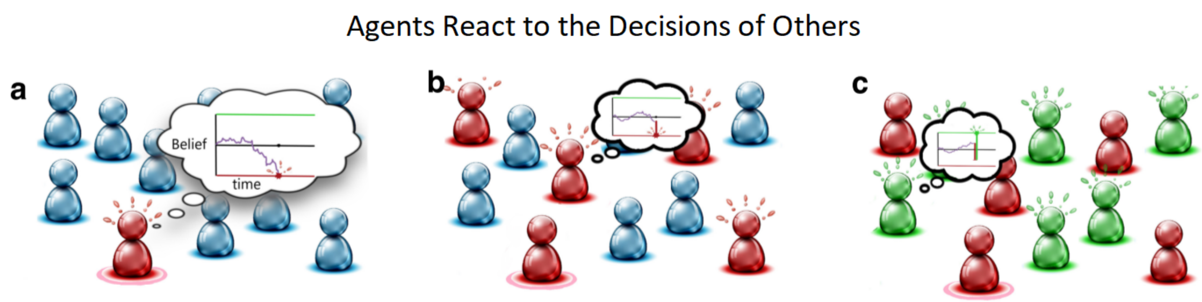

Figure 2.2: An example of updating waves in a group of agents with identical thresholds ($\theta_i = \theta$ for $1 \leq i \leq N$). (a) The first in a clique of identical agents gathers sufficient private evidence, but decides incorrectly (red). (b) The first decision provides evidence in the negative direction, convincing a few agents to agree and form the first wave. Each agent in the first wave provides evidence in the negative direction and each undecided agent provides evidence in the positive direction. Since the first wave is small, it reveals to undecided (blue) agents that the first decision was likely wrong. (c) The difference between the numbers of decided agents and undecided agents leads the remaining agents to choose correctly (green). Figure adapted from [59].

before the first decision was uninformative ($Soc(T) = 0$), we have that for agents i, $1 < i \leq N$

$$\begin{aligned} -\theta_i < y_i(T) < \theta_i \\ -\theta_i < y_{priv}^{(i)}(T) + y_{soc}^{(i)}(T) < \theta_i \\ -\theta_i < y_{priv}^{(i)}(T) < \theta_i \end{aligned}$$

for all undecided agents $j \neq 1$, $a_{j,0} = -\theta_j$ and $b_{j,0} = \theta_j$. Using these observations, we have (by symmetry)

$$\begin{aligned} Soc_j(T^{(0)}) &= \log \frac{\int_{-\theta_j}^{\theta_j} p_+^*(x,t)dx}{\int_{-\theta_j}^{\theta_j} p_+^*(x,t)dx} \\ &= 0. \end{aligned}$$

The total social evidence for the zeroth wave, then, comes from the first decider:

$$Soc(T^{(0)}) = Soc_1(T^{(0)}).$$

Since the available social evidence previous to the first decision (zeroth wave) was 0, other agents know that (supposing a first decision for H^+)

$$\begin{aligned} y_1(T) &= y^{(1)}_{priv}(T) + y^{(1)}_{soc}(T) \\ &= y^{(1)}_{priv}(T) \\ &= \theta_1. \end{aligned}$$

Moreover, since the belief of the first decider is known to be precisely at $\pm\theta_1$, rather than integrating over an interval of p^*_+ we use the known probabilities for a single agent reaching either threshold in our LLR equation:

$$\begin{aligned} Soc_1(T^{(0)}) &= \log \frac{\frac{1}{1+e^{\mp\theta_1}}}{\frac{e^{\mp\theta_1}}{1+e^{\mp\theta_1}}} \\ &= \log e^{\pm\theta_1} \\ &= \pm\theta_1. \end{aligned} \tag{2.11}$$

Hence, in the second part of the social updating step of the zeroth (first decision) wave cycle, for all agents $j \neq 1$

$$\begin{aligned} y_j(T) &= y^{(j)}_{priv}(T) + y^{(j)}_{soc}(T) \\ &= y^{(j)}_{priv}(T) \end{aligned}$$

is updated to

$$\begin{aligned} y_j(T^{(0)}) &= y^{(j)}_{priv}(T^{(0)}) + y^{(j)}_{soc}(T^{(0)}) \\ &= y^{(j)}_{priv}(T) + (Soc(T^{(0)}) - Soc_j(T^{(0)})) \\ &= y^{(j)}_{priv}(T) + (\pm\theta_1 - 0) \\ &= y^{(i)}_{priv}(T) \pm \theta_1. \end{aligned}$$

Now in step 3 of the zeroth wave cycle, we check to see whether the belief of any undecided agent has reached threshold. If it has, we move to step 1 of the first wave cycle.

During the decision step of the wave cycle, any agent still undecided considers their evidence and makes a decision if their belief has reached a threshold, joining wave A_1.

We now move to the second step of the first wave cycle, social updating. Because for all agents other than the first decider $y_{soc}^{(j)}(T^{(0)}) \neq 0$, both decisions and non-decisions are now informative. Any agent who decides in the first wave makes the same decision as the first decider: if the first decision is for $H^{\pm}$ so that $Soc(T^{(0)}) = \pm\theta_1 \gtrless 0$, social information is positive/negative and can only drive beliefs over/under the positive/negative threshold.

For $i \in A_1$, all agents know that

$$y_i(T^{(0)}) = y_{priv}^{(i)}(T) + \theta_1 > \theta_i.$$

Therefore, it may be inferred that for $i \in A_1$,

$$y_{priv}^{(i)}(T) \in [\theta_i - \theta_1, \theta_i).$$

We also know from the previous, zeroth wave that

$$y_{priv}^{(i)}(T) \in (-\theta_i, \theta_i).$$

so that $a_{i,1} = \max(-\theta_i, \theta_i - \theta_1)$, $b_{i,1} = \theta_i$ and

$$Soc_i(T^{(1)}) = \log \frac{\int_{a_{i,1}}^{b_{i,1}} p_+^*(x)dx}{\int_{-b_{i,1}}^{-a_{i,1}} p_+^*(x)dx}.$$

For $j \notin A_1$, observers know that

$$y_j(T^{(0)}) = y_{priv}^{(j)}(T) + \theta_1 < \theta_j$$

so that

$$y_{priv}^{(j)}(T) \in (-\theta_j, \theta_i - \theta_1],$$

and $a_{j,1} = -\theta_j$, $b_{j,1} = \theta_i - \theta_1$ and $Soc_j(T^{(1)})$ is calculated in the same way as $Soc_i(T^{(1)})$ above.

Because the first decider decided in the last wave, that agent provides no new information this round and

$$Soc_1(T^{(1)}) = Soc_1(T^{(0)}) = \theta_1. \tag{2.12}$$

Building on the specific examples detailed above, we may say that generally in the second step of wave cycle W, the private evidence of any agent i lies in the interval $(a_{i,W}, b_{i,W})$ whose boundaries are given with

$$a_{i,W} = \max_{1 \le m \le W} (\alpha_{i,m})$$

and

$$b_{i,W} = \min_{1 \le m \le W} (\beta_{i,m})$$

where $\alpha_{i,W}$, $\beta_{i,W}$ are given as follows:

1. Prior to the first decision, all agents' private evidence lies within the interval $(-\theta_i, \theta_i)$ but no more specific information is available: $\alpha_{i,0} = -\theta_i$ and $\beta_{i,0} = \theta_i$.

2. If agent i is a member of a previous wave ($i \in A_m$, $m < W$), then all possible information about their private evidence is already known and the boundaries of the interval in which their private evidence lies do not change: $\alpha_{i,W} = \alpha_{i,W-1}$, $\beta_{i,W} = \beta_{i,W-1}$.

3. If agent i is not a member of a previous wave ($i \notin \bigcup_m A_m$, $0 \le m < W$),

   - If $y_{soc}^{(i)}(T^{W-1}) > 0$ and $i \in A_W$: $\alpha_{i,W} = \theta_i - y_{soc}^{(i)}(T^{W-1})$, $\beta_{i,W} = \theta_i$
   - If $y_{soc}^{(i)}(T^{W-1}) > 0$ and $i \notin A_W$: $\alpha_{i,W} = -\theta_i$, $\beta_{i,W} = \theta_i - y_{soc}^{(i)}(T^{W-1})$

- If $y_{soc}^{(i)}(T^{W-1}) < 0$ and $i \in A_W$: $\alpha_{i,W} = -\theta_i$, $\beta_{i,W} = -\theta_i - y_{soc}^{(i)}(T^{W-1})$
- If $y_{soc}^{(i)}(T^{W-1}) < 0$ and $i \notin A_W$: $\alpha_{i,W} = -\theta_i - y_{soc}^{(i)}(T^{W-1})$, $\beta_{i,W} = \theta_i$

To complete the second step of the wave- triggering sequence, all undecided agents $j \notin \{A_k\}$, $0 \leq k \leq W$ then update their belief to

$$y_j(T^W) = y_{priv}^{(j)}(T) + y_{soc}^{(j)}(T^W).$$

If any agents have reached threshold using this update, another wave is triggered. With each successive wave, the bounds on the inferred position of each undecided agent's private evidence become tighter and the interval in which that private evidence might lie becomes smaller.

If social evidence is sufficiently large, waves continue until all agents have reached a decision. If the social evidence is insufficiently large, eventually there will be no agents whose belief is close enough to a threshold for social evidence to push its belief beyond threshold. In this case, according to the model, social updating would cease and the accumulation of individual evidence according to Eq. (2.2) would resume. However, social evidence garnered from indecision would subsequently change continuously in time: the waves reveal information about the amount of private evidence each undecided agent had at the time of the first decision, breaking symmetry and causing non-decisions to be continually informative (see Section 4 of Karamched *et al.*, [60]). Our analysis focuses on the behavior of the group up to the final round of social updating following the first decision and does not deal with events following the end of the first decision wave cascade.

Chapter 3

Results: Homogeneous thresholds

Here we analyse the continuous model given in the last chapter in the case where all agents have identical (*homogeneous*) thresholds: $\theta_i = \theta$ for $1 \leq i \leq N$. Until the time of the first decision, agents' private beliefs evolve according to $dy = dt + \sqrt{2}dW$, where W is a standard Wiener process. The drift and diffusion coefficients have been normalized. In this case, omniscience and consensus bias are equivalent: any agent who believes all other agents have the same threshold as themselves will be right. Material in this chapter has been published in [59].

3.1 Method of Images for calculating belief distributions

From Eq. (2.10), we have that the amount of information provided by the decision state of a single agent is given by

$$Soc_i(t) = \log \frac{\int_a^b p_+^*(x,t)dx}{\int_{-b}^{-a} p_+^*(x,t)dx}$$

where the boundaries a, b give the interval in which agent i's private information $y_{priv}^{(i)}(t)$ is known to lie.

As we noted previously, the distribution p_+^* of beliefs evolving according to Eq. (2.2) is conditioned on none of those beliefs having left the interval $(-\theta,\theta)$ previous to time t, and satisfies the Smoluchowski equation

$$\partial_t p_+^* = -\partial_x p_+^* + \partial_{xx}^2 p_+^* \tag{3.1}$$

with initial and boundary conditions

$$p_+^*(x,0) = \delta(0);$$
$$p_+^*(\pm\theta,t) = 0.$$

Ignoring the boundary conditions, and in anticipation of a method of images solution, this Smoluchowski equation has solutions of the form

$$u(x,t) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{(x-t)^2}{4t}}.$$

More generally, the Smoluchowski equation with initial condition $p_+^*(x,0) = \delta(x_0)$ has solutions of the form

$$u(x,t;x_0) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{(x-x_0-t)^2}{4t}}.$$

We consider such solutions in the context of a shifted initial boundary value problem with Dirichlet boundary conditions

$$p_+^*(x_0+\theta,t) = p_+^*(x_0-\theta,t) = 0.$$

The method of images solution from [83] for our conditioned Smoluchowski equation can be truncated for a fairly accurate approximation of the evolving probability density. The solution uses the following superposition of general solutions to obtain a specific solution given our initial

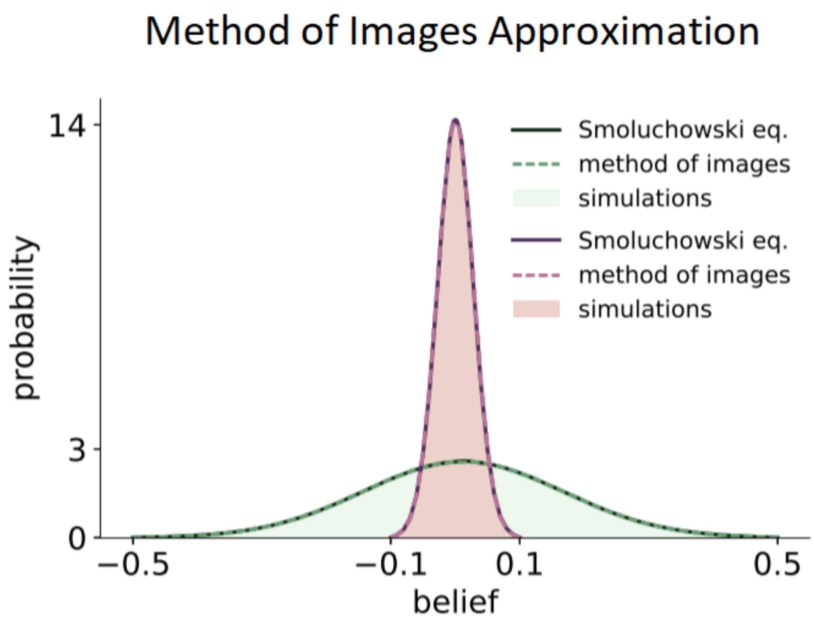

Figure 3.1: Comparison of distribution of beliefs $y_i(t)$ as described by Eq. (2.2) when computed via a numerical solution of the Smoluchowski equation, Eq. (3.1), method of images approximation given in Eq. (3.2), and empirical distribution obtained from $\mathbb{O}(10^5)$ simulations of Eq. (2.2) for $\theta = 0.1$ (red) and $\theta = 0.5$ (green). Figure from [59].

condition:

$$p_+^*(x,t) = \frac{1}{\sqrt{4\pi t}} \sum_{n=-\infty}^{\infty} \left[\exp\left(\frac{2n\theta}{4} - \frac{(x-2n\theta-t)^2}{4t} \right) - \exp\left(\frac{-2n\theta}{4} - \frac{(x+2n\theta-t)^2}{4t} \right) \right].$$

Our primary interest in this solution is at values of t reasonably close to the expected range of first decision times. When restricting to solutions on this timescale, we found that truncating the series to $n = -1, 0, 1$ provides a reasonable approximation (see Figure 3.1); hence, we have used in our analysis and computations the formula

$$p_+^*(x,t) \approx u(x,t;0) - e^{\theta} u(x,t;2\theta) - e^{-\theta} u(x,t;-2\theta). \tag{3.2}$$

3.2 Social information for agents in homogeneous cliques

In the previous chapter we noted that the available social information after the first decision is

$$Soc(T^{(0)}) = Soc_1(T^{(0)}) = \theta_1$$

where we have made, without loss of generality, the assumption that the first decider is agent 1 and that agent 1's chose the H^+ environmental state.

In the second part of the zeroth wave (see Section 2.5 for explanation of parts of each wave), all other agents then updated their belief with the new evidence $y_{soc}^{(i)}(T^{(0)}) = \pm\theta_1$ so that the evidence provided by some agent i who joins the first wave, A_1, is

$$Soc_i(T^{(1)}) = \log \frac{\int_{\max(-\theta_i, \theta_i - \theta_1)}^{\theta_i} p_+^*(x,T)dx}{\int_{-\theta_i}^{-\max(-\theta_i, \theta_i - \theta_1)} p_+^*(x,T)dx}.$$

When we have homogeneous thresholds so that $\theta_i = \theta$ for all agents $1 \leq i \leq N$, this simplifies to

$$Soc_i(T^{(1)}) = \log \frac{\int_0^{\theta} p_+^*(x,T)dx}{\int_{-\theta}^{0} p_+^*(x,T)dx}$$

for agents $i \in A_1$ in the first wave; we designate the variable $R_+(T)$ to refer to this quantity:

$$R_+ := \log \frac{\int_0^{\theta} p_+^*(x,T)dx}{\int_{-\theta}^{0} p_+^*(x,T)dx}. \tag{3.3}$$

Similarly, the amount of evidence provided by each agent who is still undecided after the first wave, $j \notin A_0 \bigcup A_1$, is given by

$$
\begin{aligned}
Soc_j(T^{(1)}) &= \log \frac{\int_{-\theta}^{0} p_+^*(x,T)dx}{\int_0^{\theta} p_+^*(x,T)dx} \\
&= -R_+(T)
\end{aligned}
$$

Due to the homogeneity of the thresholds, all undecided agents provide the same amount of information, and all agents in the first wave provide the same amount of information. Moreover, the amount of evidence provided by an agent in the first wave is equal in magnitude and opposite in sign to the evidence provided by an agent still undecided after the first wave.

Let a_W be the number of agents in a wave A_W and u_W the number of agents still undecided after that same wave. These quantities are related by

$$
u_W = N - \sum_{k=0}^{W} a_k.
$$

Accordingly, the amount of social evidence each undecided agent is provided by the first wave (excluding the $\pm\theta$ provided by the zeroth wave whose sole member is the first decider, agent 1), $y_{soc}^{(j)}(T^{(1)})$, is given by

$$
\begin{aligned}
y_{soc}^{(j)}(T^{(1)}) &= Soc(T^{(1)}) - Soc_j(T^{(1)}) - Soc_1(T^{(1)}) \\
&= \sum_{2}^{N} Soc_i(T^{(1)}) - Soc_j(T^{(1)}) \\
&= \sum_{2}^{a_1} R_+(T) + \sum_{2}^{u_1}(-R_+(T)) - (-R_+(T)) \\
&= a_1 R_+(T) + (u_1 - 1)(-R_+(T)) \\
&= a_1(R_+(T)) + (N - (a_1 + a_0) - 1)(-R_+(T)) \\
&= R_+(T)(2a_1 - N + 2).
\end{aligned}
$$

We used the fact that the zeroth wave contains only the first decider, so that a_0 will always be 1. We let c_1^+ refer to the social increment $y_{soc}^{(j)}(T^{(1)})$ all undecided agents j receive after the first wave when the first decision is H^+. If the first decision is H^-, similar calculations give us the quantity $c_1^- = -R_+(T)(2a_1 - N + 2)$ as the social increment after the first wave. We will use the terms *social increment* and *social update* interchangeably.

Since the social evidence, $y_{soc}^{(i)}(T^{(0)})$, is equal in sign to the direction of the first decision, any agents deciding in the first wave will make the same decision as the first decider. The second wave is more variable: If $a_1 > N/2 - 1$, then c_1^+ will have the same sign as the first decision, providing evidence in favor of H^+, so that any agents who decide in the second wave will agree with the decision of the first decider. If $a_1 < N/2 - 1$, the evidence c_1^+ will disagree with the decision of the first decider and any agents deciding in the second wave will disagree with the first decider. We will show that for sufficiently large N, the entire group is likely to have decided by the end of the second wave.

3.3 First decision time

Both the value of R_+ (the social increment resulting from the first wave) and the expected size of our first wave depend on the time of the first decision. We therefore next compute the expectation of the first decision time. Let $\rho_\pm(x,t)$ be the first passage time distribution for a single agent through the absorbing boundaries $x = \pm\theta$. A good approximation of this distribution can be obtained by taking the time derivative of our method of images solution for the pdf of that agent's belief and evaluating it at those boundaries:

$$\rho_\pm(t) = \mp\frac{\partial \rho}{\partial x}|_{x=\pm\theta} = \frac{\theta}{\sqrt{4\pi t^3}} \exp\left(-\frac{(\theta \mp t)^2}{4t}\right).$$

Then our survival probabilities are given by the cdfs $\Phi_{\pm}(t)$:

$$\Phi_{\pm}(t) = \int_0^t \rho_{\pm}(s)ds = \frac{1}{2}\left[\operatorname{erfc}\frac{\theta \mp t}{2\sqrt{t}} + e^{\pm\theta}\operatorname{erfc}\frac{\theta \pm t}{2\sqrt{t}}\right].$$

The combined probabilities at both boundaries are $\rho(t) = \rho_+(t) + \rho_-(t)$ and $\Phi = \Phi_+ + \Phi_-$. Hence, for a single agent, the probability that the first passage time τ is greater than t is

$$P(\tau > t) = 1 - (\Phi_+(t) + \Phi_-(t)).$$

For a group of N agents, the probability that the first passage time in that group τ_N is smaller than t is

$$\begin{aligned} P(\tau_N > t) &= P(\tau_N > t)^N \\ &= (1 - \Phi))^N \end{aligned}$$

where $\Phi = \Phi_+ + \Phi_-$. Differentiating with respect to time gives us the distribution for the first decision time in a group of N agents:

$$p_N(t) = N\Big(1 = \Phi(t)\Big)^{N-1}\Big(\rho(t)\Big)$$

which may be rewritten as

$$p_N = N\exp\left((N-1)\ln(1-\Phi(t))\right)\rho(t).$$

If t is small we can use the Taylor expansion of the last expression to obtain

$$p_N(t) \approx Ne^{-(N-1)\Phi(t)}\rho(t). \tag{3.4}$$

We see the agreement of Eq. (3.4) with simulation results in Figure 3.2.

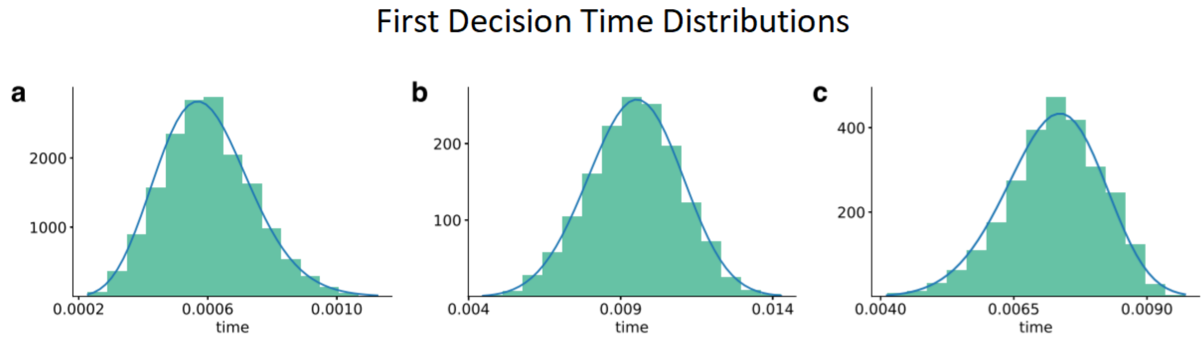

Figure 3.2: First decision time distributions obtained numerical simulations (green bars) and theory (blue curves) from Eq. (3.4). (a) $N = 100$ and $\theta = 0.1$; (b) $N = 1200$ and $\theta = 0.5$; (c) $N = 10000$ and $\theta = 0.5$. Figure from supplemental material for [59].

Returning to the first passage time for a single agent, we had

$$\begin{aligned} P(\tau > t) &= 1 - (\Phi_+(t) + \Phi_-(t)) \\ &= \frac{1}{2}\left[\operatorname{erf}\frac{\theta + t}{2\sqrt{t}} + \operatorname{erf}\frac{\theta - t}{2\sqrt{t}} - e^{\theta}\operatorname{erfc}\frac{\theta + t}{2\sqrt{t}} - e^{-\theta}\operatorname{erfc}\frac{\theta - t}{2\sqrt{t}}\right]. \end{aligned}$$

For t small, we can approximate this expression by

$$P(\tau > t) \approx \frac{1}{2}\left[\operatorname{erf}\frac{\theta + t}{2\sqrt{t}} + \operatorname{erf}\frac{\theta - t}{2\sqrt{t}}\right].$$

To extend the first passage time calculation to accommodate N agents, we want to obtain the probability that our smallest passage time, τ_N, is greater than t:

$$\begin{aligned} P(\tau_N > t) &= P(\tau_N > t)^N \\ &\approx \frac{1}{2^n}\left[\operatorname{erf}\frac{\theta + t}{2\sqrt{t}} + \operatorname{erf}\frac{\theta - t}{2\sqrt{t}}\right]^N \\ &\approx \left(\operatorname{erf}\frac{\theta}{2\sqrt{t}}\right)^N, \end{aligned}$$

where the second approximation also uses the assumption that t is small. Then we can describe the expectation for the smallest passage time (first decision time) τ_N with

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}[\tau_N] &= \lim_{N\to\infty}\int_0^\infty P(\tau_N > t)dt \\ &\approx \lim_{N\to\infty}\int_0^\infty \left(\operatorname{erf}\frac{\theta}{2\sqrt{t}}\right)^N dt \\ &\approx \lim_{N\to\infty}\int_0^\infty \exp\left(-\frac{2N}{\theta}\sqrt{\frac{t}{\pi}}e^{-\frac{\theta^2}{4t}}\right)dt = 0.\end{aligned}$$

By extreme value theory [41] there exists an asymptotic scaling $b_n \to 0$ such that

$$\lim_{N\to\infty} P(\tau_N > b_N t) = \lim_{N\to\infty}\operatorname{erf}\left(\frac{\theta}{\sqrt{2b_N t}}\right)^N = F(t)$$

where $F(t)$ is the cdf for the Gumbel distribution.

To obtain the sequence b_N, we define a sequence t_N where $P(\tau_N > t_N) = p \in (0,1)$:

$$t_N = \frac{\theta^2}{4}\left(\operatorname{erf}^{-1}(p^{1/N})\right)^{-2}. \tag{3.5}$$

The error function has an asymptotic expansion

$$\left(\operatorname{erf}^{-1}(x)\right)^2 \approx -\ln(\sqrt{\pi}(1-x))$$

that holds for $x \approx 1$. Substituting this expansion into Eq. 3.5, we get

$$\frac{\theta^2}{4}\frac{1}{t_N} \approx -\ln\left(\sqrt{\pi}(1-p^{1/N})\right).$$

Using the further approximation $(1-p^{1/N}) \approx \ln(p^{1/N})$,

$$\frac{\exp(\frac{\theta^2}{2t_N})}{N^2} \approx \frac{1}{\pi(\ln p)^2}.$$

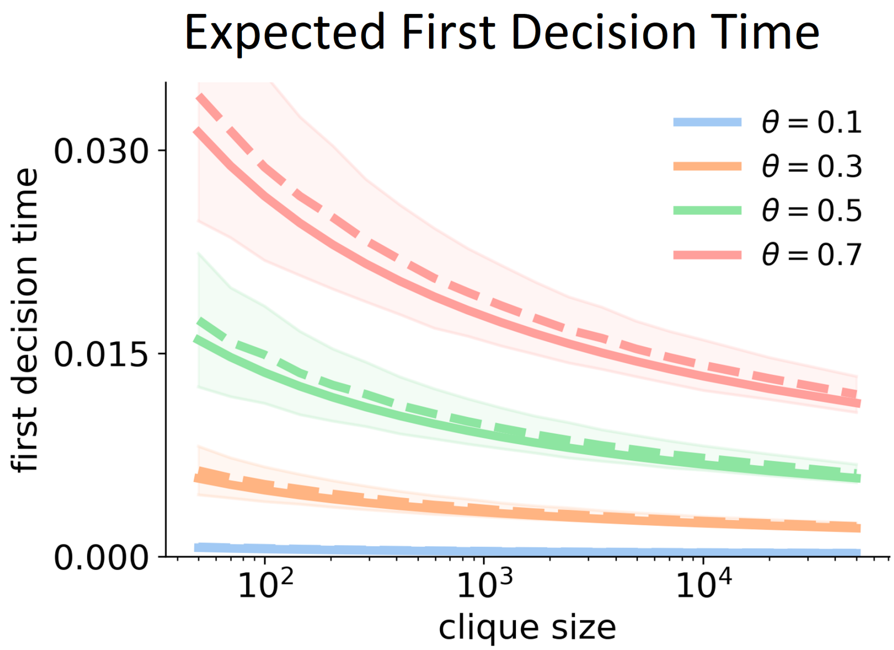

Figure 3.3: Expected first decision time as a function of clique size for homogeneous threshold cliques with various threshold values. Here, and below, solid and dashed lines represent simulations and theory, respectively, and shaded regions capture one standard deviation of simulated results around the mean. The solid lines give Eq. (3.6). Figure from supplemental material for [59].

We then see that the expected first decision time can be approximated by

$$\mathbb{E}[T] = \mathbb{E}[\tau_N] \approx \frac{\theta^2}{2\ln N^2} = \frac{\theta^2}{4\ln N}. \tag{3.6}$$

Figure 3.3 compares Eq. (3.6) with simulation results.

So as N increases, our expected first decision time decreases logarithmically with it and for very large N, our first decision time T is very small. We will use our expectation for T heavily in later calculations.

3.4 Expected size of the first wave

The expected size of the first wave following a correct first decision made at time T is the probability that a given agent's belief will be in the upper half of the symmetric interval $(-\theta, \theta)$ at the time of the first decision conditioned on their belief having never left that interval multiplied by the number of agents remaining after the first decider has been removed:

$$\mathbb{E}[a_1|T] = (N-1)\frac{\int_0^\theta p_+^*(x,T)dx}{\int_{-\theta}^\theta p_+^*(x,T)dx}.$$

If we approximate $p_+^*(x,t)$ using only the first term of our method of images solution and retain the assumption that our first decision time T is small, we have

$$\begin{aligned}\mathbb{E}[a_1|T] &\approx (N-1)\frac{\operatorname{erf}\frac{\theta-T}{2\sqrt{T}} + \operatorname{erf}\frac{\sqrt{T}}{2}}{\operatorname{erf}\frac{\theta-T}{2\sqrt{T}} + \operatorname{erf}\frac{\theta+T}{2\sqrt{T}}} \\ &= \frac{N-1}{2}\left(1 + \frac{\operatorname{erf}\frac{\sqrt{T}}{2}}{\operatorname{erf}\frac{\theta}{2\sqrt{T}}}\right).\end{aligned}$$

After using the Taylor expansion for $\operatorname{erf}\frac{\sqrt{T}}{2}$ and the small T asymptotic expansion for $\operatorname{erf}\frac{\theta}{2\sqrt{T}}$ this becomes

$$\mathbb{E}[a_1|T] \approx \frac{N-1}{2}\left(1 + \frac{\sqrt{T/\pi}}{1 - \frac{2}{\theta}\sqrt{\frac{T}{\pi}}e^{-\frac{\theta^2}{4T}}}\right).$$

If we then substitute in our expected first decision time $\mathbb{E}[T] \approx \frac{\theta^2}{4\ln N}$, we get our expected first wave size

$$\mathbb{E}[a_1] \approx \frac{N-1}{2}\left(1 + \frac{\theta/2}{\sqrt{\pi \ln N} - \frac{1}{N}}\right).$$

Assuming further that N is large lets us reduce this to

$$\mathbb{E}[a_1] \approx \frac{N-1}{2}\left(1+\frac{\theta}{\sqrt{4\pi \ln N}}\right) \tag{3.7}$$

for first waves following correct first decisions.

In the case of an incorrect first decision, we begin with

$$\mathbb{E}[a_1|T] = (N-1)\frac{\int_{-\theta}^{0} p_+^*(x,T)dx}{\int_{-\theta}^{\theta} p_+^*(x,T)dx}$$

and follow a similar derivation to obtain

$$\mathbb{E}[a_1] \approx \frac{N-1}{2}\left(1-\frac{\theta}{\sqrt{4\pi \ln N}}\right).$$

Thus, we have that in the case of a correct first decision, slightly more than half the clique will, on average, follow and chose rightly in the first wave and in the case of an incorrect first decision, slightly less than half the clique, on average, will follow and choose wrongly in the first wave. As N increases, we note that while the fraction of agents deciding in the first wave approaches one half, the absolute difference between the number of deciding and non-deciding agents after the first wave continues to grow. (See Figure 3.4.)

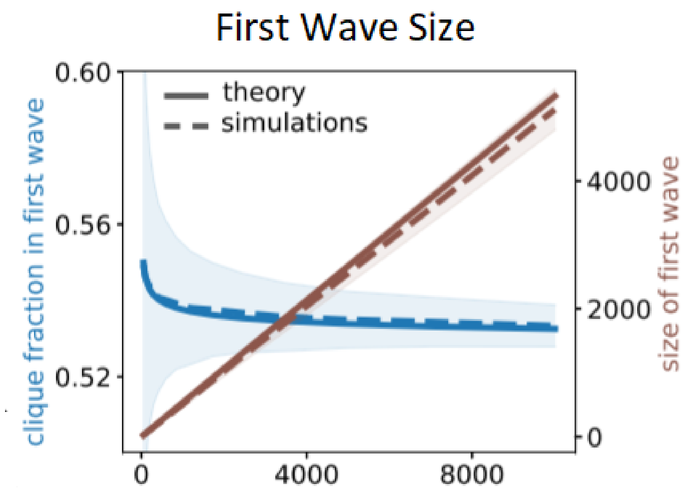

Figure 3.4: The first wave increases with N (red), but comprises a smaller fraction of the population (blue; Eq. (3.7)). Here, the first decision is correct and $\theta = 0.7$. Taken from [59].

3.5 Expected amount of information provided by an agent deciding in the first wave

We recall from Section 3.2 that in the case of a correct first decision, an agent deciding in the first wave makes available social information in the amount

$$R_+(T) = \log \frac{\int_0^\theta p_+^*(x,T)dx}{\int_{-\theta}^0 p_+^*(x,T)dx}$$

and an agent still undecided after the first wave made available social information in the amount of

$$-R_+(T) = \log \frac{\int_{-\theta}^0 p_+^*(x,T)dx}{\int_0^\theta p_+^*(x,T)dx}.$$

In the case of an incorrect first decision, these quantities would be reversed. We would then like to obtain the expected size of this information R_+ conditioned on the time of the first decision. First we note that

$$
\begin{aligned}
R_+(T) &= \log\left(\frac{\int_0^\theta p_+^*(x,T)dx}{\int_{-\theta}^\theta p_+^*(x,T)dx}\right) \\
&= \log\left(\int_0^\theta p_+^*(x,T)dx\right) - \log\left(\int_{-\theta}^\theta p_+^*(x,T)dx\right) \\
&\approx \log\left(\operatorname{erf}\frac{\theta}{2\sqrt{T}} + \operatorname{erf}\frac{\sqrt{T}}{2}\right) - \log\left(\operatorname{erf}\frac{\theta}{2\sqrt{T}} - \operatorname{erf}\frac{\sqrt{T}}{2}\right) \\
&\approx \log\left(1 - \frac{2}{\theta}\sqrt{\frac{T}{\pi}}e - \frac{\theta^2}{4T} + \sqrt{\frac{T}{\pi}}\right) - \log\left(1 - \frac{2}{\theta}\sqrt{\frac{T}{\pi}}e - \frac{\theta^2}{4T} - \sqrt{\frac{T}{\pi}}\right) \\
&\approx -\frac{2}{\theta}\sqrt{\frac{T}{\pi}}e - \frac{\theta^2}{4T} + \sqrt{\frac{T}{\pi}} + \frac{2}{\theta}\sqrt{\frac{T}{\pi}}e - \frac{\theta^2}{4T} + \sqrt{\frac{T}{\pi}} \\
&= 2\sqrt{\frac{T}{\pi}}.
\end{aligned}
$$

Then taking the expectation,

$$\mathbb{E}[R_+(T)|T] \approx 2\mathbb{E}[\sqrt{\frac{T}{\pi}}].$$

Substituting in our expected value for T yields

$$\mathbb{E}[R_+(T)] \approx \frac{\theta}{\sqrt{\pi \ln N}}. \tag{3.8}$$

We can see this approximation compared to simulation results in Figure 3.5. We note that as our clique size N increases and our first decision time T decreases along with it, the expected value of R_+ decreases as well. This is because for very small decision times, the beliefs of the majority of agents are still clustered close to origin with a nearly symmetric distribution; the ratio of the area of the distribution p_+^* above 0 to the area of the distribution below 0 is very close to 1.

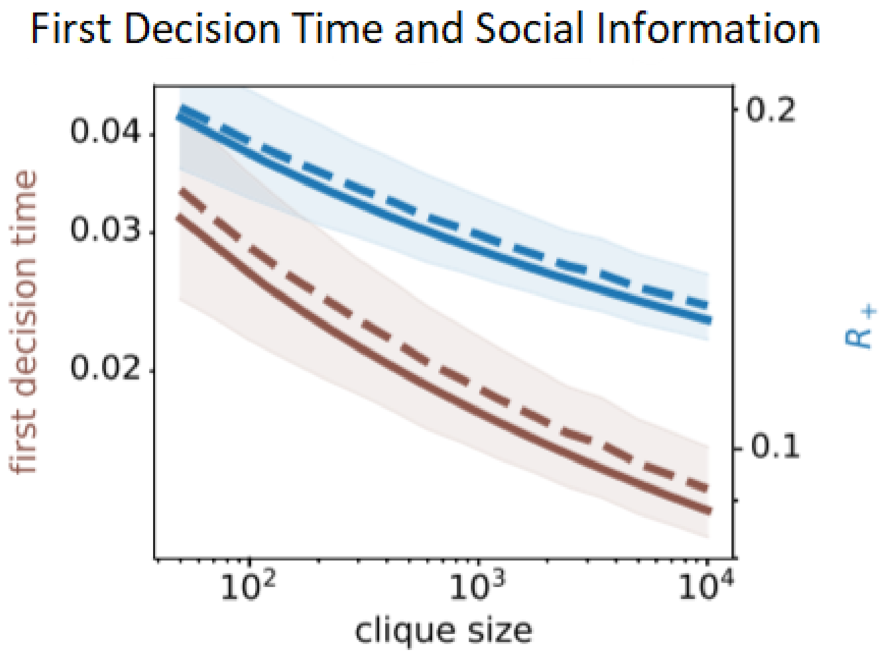

Figure 3.5: The time to the first decision decreases with network size (brown), allowing each agent less time to accumulate private information (Eq. (3.6)). Information provided by an individual deciding in the first wave also decreases ($R+$, blue; Eq. (3.8)). $\theta = 0.7$. Figure from [59].

3.6 Finding the expected social increment after first wave, $\hat{c}_1^{\pm}$

Earlier we asserted that for sufficiently large groups, the entire clique will have decided by the end of the second wave. To show this, we begin by finding the relationship between the group size N and the amount of social information undecided agents receive after the first wave. In Section 3.2 we introduced this quantity as c_1^+, the size of the social update (increment) each undecided agent receives after the first wave following a positive first decision. We know that an agent j who is undecided after the first wave must have private evidence in the interval $(-\theta, 0]$. Consequently, if $c_1^+ >= 2\theta$, agent j's total evidence at the beginning of the second wave cycle will be larger than θ and agent j will join the second wave.

From Section 3.2 we have

$$c_1^+ = R_+(T)(2a_1 - N + 2),$$

so that c_1^+ depends both on the amount of information R_+ provided by a single agent and on the total number of agents joining the first wave. The size of the first wave grows linearly in N (Eq. (3.6)) but this growth is countered by a concurrent logarithmic decrease in R_+ (Eq. (3.8)).

We seek an expectation $\hat{c}_1^+$ for the size of c_1^+. In the case of a correct first decision,

$$\begin{aligned}
\mathbb{E}[c_1^+] &= 2\mathbb{E}[a_1]\mathbb{E}[R_+] - (N-2)\mathbb{E}[R_+] \\
&= 2\frac{N-1}{2}\left(1 + \frac{\theta}{\sqrt{4\pi \ln N}}\right)\frac{\theta}{\sqrt{\pi \ln N}} - (N-2)\frac{\theta}{\sqrt{\pi \ln N}} \\
&= \frac{\theta}{\sqrt{\pi \ln N}} + \frac{\theta^2 (N-1)}{2\pi \ln N},
\end{aligned}$$

which can be approximated for large N by

$$\mathbb{E}[c_1^+] \approx \frac{\theta^2 N}{2\pi \ln N}.$$

Similarly, when the first decision is wrong we have

$$\begin{aligned}
\mathbb{E}[c_1^-] &= (N-2)\frac{\theta}{\sqrt{\pi \ln N}} - \frac{N-1}{2}\left(1 - \frac{\theta}{\sqrt{4\pi \ln N}}\right)\frac{\theta}{\sqrt{\pi \ln N}} \\
&= -\frac{\theta}{\sqrt{\pi \ln N}} + \frac{\theta^2 (N-1)}{2\pi \ln N} \\
&\approx \frac{\theta^2 N}{2\pi \ln N}.
\end{aligned}$$

Thus, the expected size of the new social information made available as a result of agents deciding or not deciding in the first wave is the same and positive, irrespective of the accuracy of the first

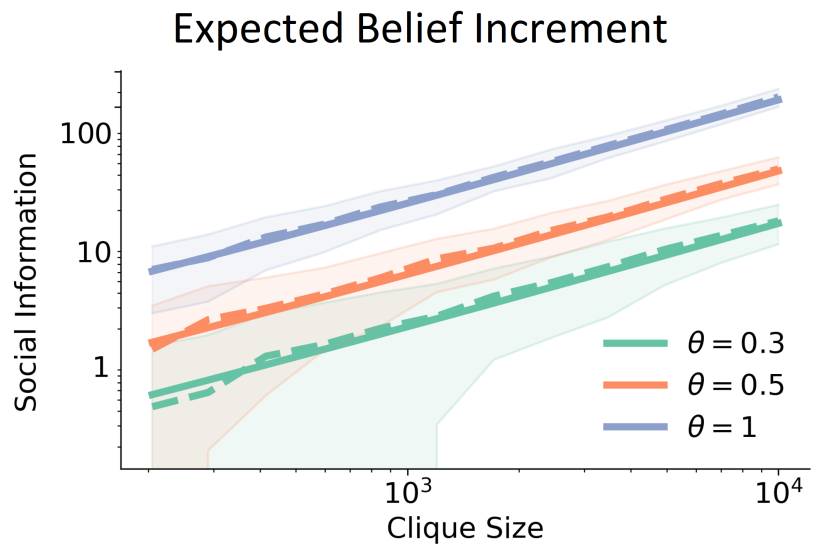

Figure 3.6: Expected belief increment after the first wave due to social information $\hat{c}_1^{\pm}$ as given by Eq. (3.9). Figure taken from [59].

decider:

$$\hat{c}_1^{\pm} := \frac{\theta^2 N}{2\pi \ln N} \approx \mathbb{E}[c_1^{+}] \approx \mathbb{E}[c_1^{-}]. \tag{3.9}$$

Figure 3.6 compares Eq. (3.9) If the first decision is correct, more than half the network is in the first wave, and both $(2\mathbb{E}[a_1] - N - 2)$ and $R_+(T)$ are positive. Both of these terms are negative when the first decision is wrong. Thus the second wave is self-correcting: in large networks, even if the first decision and first wave are wrong, all undecided agents make the correct choice in the second wave. When the network is sufficiently large, $\hat{c}_1^{\pm} > 2\theta$. However, in many trials the actual increment to social information will fall below the expected value. To estimate the probability that $c_1^{\pm} > 2\theta$ in a network of size N, we next provide a lower bound on the variance of the increment $c_1^{\pm}$.

3.7 Variance of social increment $c_1^{\pm}$ and its lower bound

The increment $c_1^{\pm}$ depends on the size of the first wave; hence, we begin by finding the variance of the size of the first wave, $\mathbb{V}[a_1]$. Let $p_A(t)$ be the probability that a lone agent who is undecided at time t has a private belief that satisfies $y_{priv}^{(i)}(t) \geq 0$:

$$p_A(t) = \frac{\int_0^{\theta} p_+^*(x,t)dx}{\int_{-\theta}^{\theta} p_+^*(x,t)dx}$$

Then we may describe the size of the first wave with a binomial distribution:

$$P[a_1 = n] = \binom{N-1}{n} p_A(t)^n \Big(1 - p_A(t)\Big)^{N-n-1}$$

so that

$$\mathbb{V}[a_1] = (N-1)p_A(t)\Big(1 - p_A(t)\Big). \tag{3.10}$$

and

$$\begin{aligned} \mathbb{V}[c_1^{\pm}] &= R_+(T)^2 \mathbb{V}[a_1] \\ &\approx \frac{4\theta^2}{\pi}\frac{N-1}{\ln N}\left(1 - \frac{\theta^2}{4\pi \ln N}\right). \end{aligned} \tag{3.11}$$

Figure 3.7 compares Eq. (3.10) and Eq. (3.11) to simulation results.

To obtain a lower bound on this variance of $c_1^{\pm}$, we use Chebyshev's inequality which states that for a random variable X with finite mean μ and non-zero variance σ^2, and for all $k \in \mathbb{R}^+$,

$$P(|X - \mu| \geq \sigma k) \leq \frac{1}{k^2}.$$

We desire a lower bound on N to guarantee that with probability x the entire clique will decide

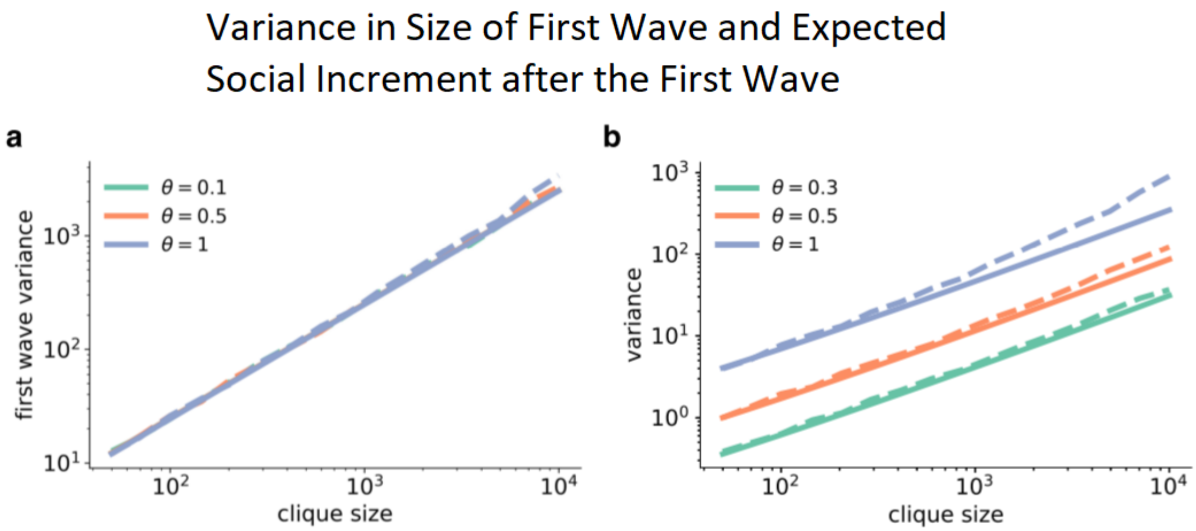

Figure 3.7: (a) Variance in the size of the first wave a_1 from simulations (dashed) and theory Eq. (3.10) (solid) for various θ. (b) Variance in the expected social increment after the first wave $c_1^{\pm}$ from simulations and Eq. (3.11) for various thresholds θ. Figure taken from supplemental material for [59].

by the end of the second wave. Letting $k = (1-x)^{-\frac{1}{2}}$, this lower bound has the form

$$\frac{|c_1^{\pm} - \mathbb{E}[c_1^{\pm}]|}{\sigma} \geq k$$

where $\sigma^2 = \mathbb{B}[c_1^{\pm}]$. Since we wish to ensure $c_1^{\pm} \geq 2\theta$, this becomes

$$\frac{|2\theta - \mathbb{E}[c_1^{\pm}]|}{\sigma} \geq \frac{1}{\sqrt{1-x}}.$$

After substituting in our values for the expectation and variance of $c_1^{\pm}$, we have

$$\frac{\frac{N\theta^2}{2\pi \ln N} - 2\theta}{\sqrt{\frac{\theta^2(N-1)}{\pi \ln N}}\sqrt{1 - \frac{\theta^2}{4\pi \ln N}}} \approx \frac{\frac{N\theta^2}{2\pi \ln N}}{\sqrt{\frac{N}{\pi \ln N}}} \geq \frac{1}{\sqrt{1-x}}$$

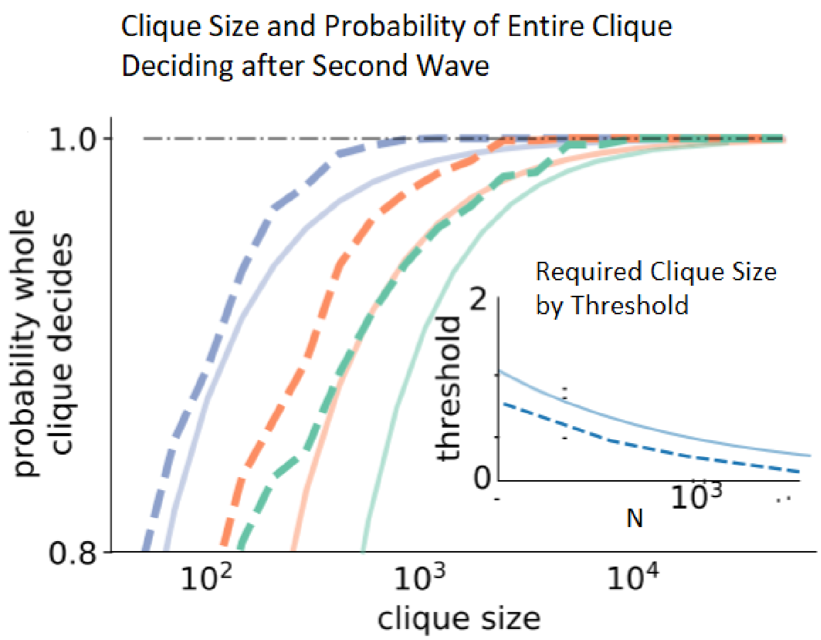

Figure 3.8: Probability the full clique decides after the second wave. Chebyshev's Inequality provides a lower bound on clique size N by which the probability is reached (Eq. (3.12)). Inset: Threshold θ at which expected social increment after the first wave $\hat{c}_1^{\pm} = 2\theta$, the maximum size of the interval in which undecided agents' beliefs lie, as clique size N varies on the horizontal axis. Figure taken from [59].

and

$$\frac{N\theta^2}{4\pi \ln N} \geq \frac{1}{\sqrt{1-x}}.$$

Since

$$\frac{N\theta^2}{4\pi} > \frac{N\theta^2}{4\pi \ln N},$$

we may then give our bound as

$$N \geq \frac{4\pi}{\theta^2(1-x)} \tag{3.12}$$

to guarantee the entire clique will decide by the end of the second wave with probability x or greater. Figure 3.8 compares Eq. (3.12) to simulation results.

3.8 Expected probability of a random agent deciding correctly

We now wish to consider the probability that any randomly chosen agent i in a homogeneous-threshold clique will have made a correct decision rather than deciding incorrectly or remaining undecided after all waves following the first decision have concluded. We will abuse notation and define $P^{+}(y_i(T^W) \geq \theta_i)$ as the probability that agent i decides correctly in some wave $0 \leq W$. This probability is given by the fraction of the group that has decided accurately after all wave cycles have concluded.

In the previous section (Section 3.7) we showed that in a sufficiently large group, we expect that all agents will have decided by the end of the second wave. When this happens, we expect that the decisions will follow one of two patterns:

1. The first decider is correct, the first wave is correct, and the second wave is correct for an expected accuracy $\mathbb{E}[P^{+}\Big(y_i(T^W) \geq \theta | y_1(T) = \theta\Big)] = 1$ where without loss of generality agent 1 makes the first decision; or

2. The first decider is incorrect and the first wave is incorrect, but the second wave is correct for an expected accuracy of $\mathbb{E}[P^{+}\Big(y_i(T^W) \geq \theta | y_1(T) = \theta\Big)] = \frac{\mathbb{E}[a_2]}{N} = \frac{N-1-\mathbb{E}[a_1]}{N}$ where a_2 is the size of the second wave.

Then without conditioning on the accuracy of the first decider, our expected accuracy for a

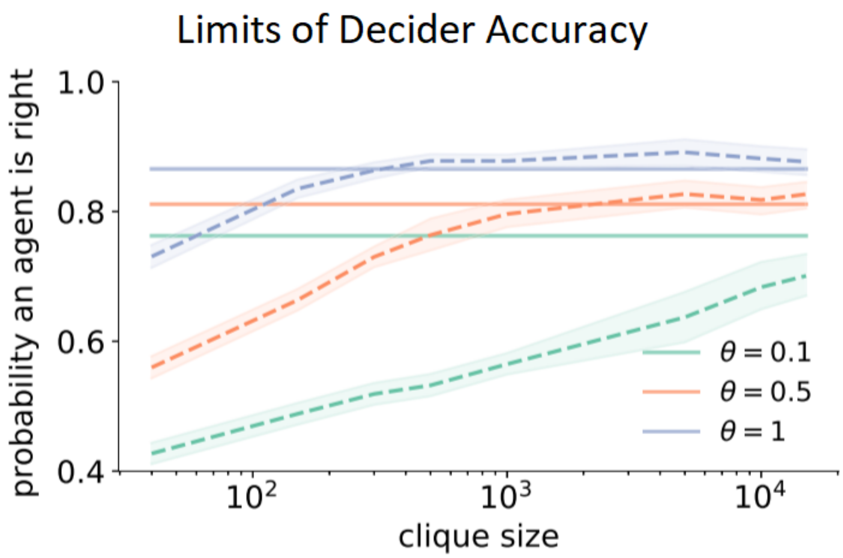

Figure 3.9: The probability that an agent in a clique will make a correct choice as a function of the size of the clique, N. This probability converges to a constant value given in Eq. (3.14). Taken from supplemental material for [59].

random agent i is

$$\begin{aligned}
\mathbb{E}[P^{+}\Big(y_i(T^W) \geq \theta\Big)] &= P^{+}\Big(y_1(T) = \theta\Big)\mathbb{E}[P^{+}\Big(y_i(T^W) \geq \theta_{|}y_1(T) = \theta\Big)] \\
&\quad + P^{+}\Big(y_1(T) = -\theta\Big)\mathbb{E}[P^{+}\Big(y_i(T^W) \geq \theta|y_1(T) = \theta\Big)] \\
&= \frac{1}{1+e^{-\theta}} + \frac{e^{-\theta}}{1+e^{-\theta}}\frac{N-1-\mathbb{E}[a_1]}{N} \\
&= \frac{1}{1+e^{-\theta}} + \frac{e^{-\theta}}{1+e^{-\theta}}\frac{1}{N}\left[N-1-\left(\frac{N-1}{2}\left(1-\frac{\theta}{\sqrt{4\pi \ln N}}\right)\right)\right].
\end{aligned} \tag{3.13}$$

As N approaches infinity, we have

$$\lim_{N\to\infty} \mathbb{E}[P^{+}\Big(y_i(T^W) \geq \theta\Big)] = \frac{1}{1+e^{-\theta}} + \frac{1}{2}\left(\frac{e^{-\theta}}{1+e^{-\theta}}\right). \tag{3.14}$$

This limited accuracy is shown in Figure 3.9 for various values of θ. The smaller the value of θ, the larger the value of N required to reach this limit. Thus, in homogeneous cliques, adding more agents can result in faster and more accurate decisions than those made by a single agent; however, there is a limit to the improvement in accuracy. At a certain point, adding more agents to the group does nothing further to improve any individual agent's chances of making a correct decision. In the next chapter, we explore whether heterogeneous groups (specifically groups with heterogeneous thresholds) labor under the same limitations.

Chapter 4

Results: Heterogeneous thresholds

In the previous chapter we showed that in cliques with homogeneous thresholds the expected proportion of the clique choosing accurately (and therefore the probability of any single agent within the clique choosing accurately) could be given as a function of the threshold θ and clique size N. For large N, this probability could be given as a function of θ. For groups sufficiently large that we might expect the entire clique to chose by the end of the second wave, we had both that the expected decision time (Eq. 3.6) and the expected accuracy (Eq. (3.13)) of any single agent in a homogeneous threshold group were faster and more accurate than those of a single agent (Eqs. (2.3) and (2.4), respectively.)

In biological situations, decision-making populations are rarely homogeneous. Some agents may decide impulsively while others require substantial evidence to make a decision. [21, 57, 101, 88] To model this diversity, we first examine a dichotomous threshold case where some fraction γ of the population has threshold θ_{min} and the remainder of the population has threshold θ_{max}. Agents with lower thresholds will on average decide more quickly but are also more likely to make a wrong choice. The ensuing exchange of social information depends on assumptions agents make about each other. If the population operates under a *consensus bias*, agents perform non-Bayesian social updating by assuming that all other agents share their own threshold. In this case, the group behaves much the same as a group with homogeneous thresholds. In an *omniscient* population, agents know all other agents' thresholds exactly and social updating is therefore rational (Bayesian). In such groups, omniscient agents can leverage quick, unreliable decisions to improve the response of the population. The material in this chapter was previously published in [59].

In both the omniscient and the consensus bias case, prior to a first decision agents' beliefs evolve according to Eq. 2.2 with absorbing boundaries at $-\theta_i < 0 < \theta_i$. A small threshold subgroup of γN agents share threshold θ_{min} and a large threshold subgroup of $(1-\gamma)N$ agents share threshold θ_{max} for $0 < \theta_{\text{min}} < \theta_{\text{max}}$ and $\gamma \in (0,1)$.

4.1 First decision time

Let $p_{N,d}(t)$ be the first passage time density for N agents whose thresholds follow a dichotomous distribution. Then the expected first decision time is

$$\mathbb{E}[T] = \int_0^\infty t\, p_{N,d}(t)\,dt.$$

Computing this integral explicitly is, in general, not possible. Instead, we approximate the expectation by assuming that the first decision is made by an agent with threshold θ_{min} and compute this approximate first decision time using methods similar to those in Section 3.3. The expected first decision time T_{min} in a group of γN agents with identical thresholds θ_{min} is

$$\mathbb{E}[T_{\text{min}} | \text{decider had threshold } \theta_{\text{min}}] \approx \frac{\theta_{\text{min}}^2}{4 \ln \gamma N}$$

and the expected first decision time $T_{\max}$ in a group of $(1-\gamma)N$ agents with identical thresholds $\theta_{\max}$ is

$$\mathbb{E}[T_{\max}|\text{decider had threshold } \theta_{\max}] \approx \frac{\theta_{\max}^2}{4\ln(1-\gamma)N}.$$

Let $P_{k,\pm}$ be the probability that the first decider has threshold k and decides correctly $(+)$ or incorrectly $(-)$. The overall expected first decision time is then given by

$$\mathbb{E}[T] \approx \Big(P_{\min,+} + P_{\min,-}\Big)\frac{\theta_{min}^2}{4\ln\gamma N} + \Big(P_{\max,+} + P_{\max,-}\Big)\frac{\theta_{\max}^2}{4\ln(1-\gamma)N}. \tag{4.1}$$

To compute $P_{k,\pm}$, let $P_{N,\pm}(t,\theta_k)\Delta t$ denote the joint probability that the first decider amongst N agents has threshold θ_k for $k=$ max, min and makes the correct $(+)$ or incorrect $(-)$ decision. The probability of a correct first decision is then $P_{k,\pm} \equiv \int_0^\infty P_{N,\pm}(t,\theta_k)dt$. Analytic representations of these probabilities are not available so we use Monte Carlo methods to evaluate them. Hence, the weights $P_{k,+} + P_{k,-} \equiv P_k$ are the probabilities that the first decider has threshold k. In general, $P_{\min} \gg P_{\max}$ so we use the approximation

$$\mathbb{E}[T] \approx \frac{\theta_{\min}^2}{4\ln\gamma N}. \tag{4.2}$$

We compare this approximation to simulations in the first graph of Figure 4.1. The approximation breaks down when $0 < \gamma \ll 1$, but works well otherwise.

4.2 Expected size of first wave under consensus bias

In groups operating under consensus bias, each agent assumes that all other agents share their same threshold: each agent i believes they are in a homogeneous clique where every agent has threshold

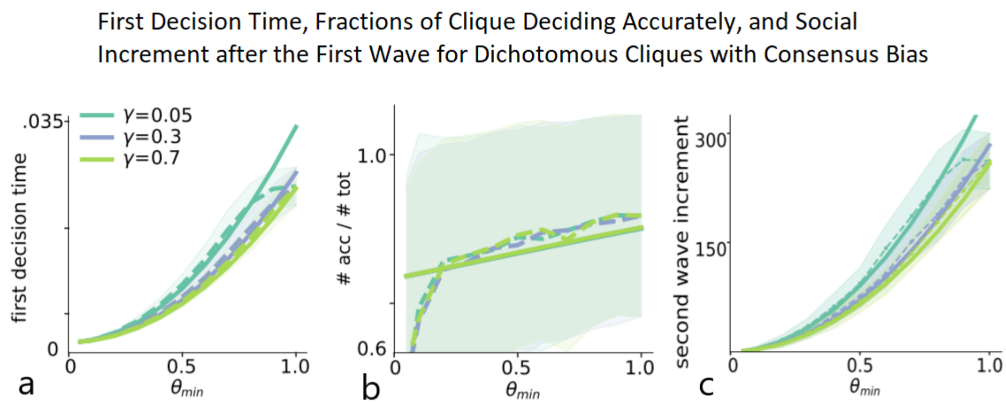

Figure 4.1: (a) First decision time for dichotomous threshold cliques for various clique fractions γ at lower threshold. (Eq. (4.2)). (b) Fraction of accurate deciders in dichotomous threshold cliques under consensus bias (Eq. (4.5)). (c) Belief increment of agents in the second wave in dichotomous threshold cliques under consensus bias (Eq. (4.4)). Clique size $N = 15000$ in panels (c–e) Figure taken from [59].

θ_i.

Therefore, when the first agent decides, all other agents in the group update their belief by their own threshold rather than by the threshold of the first decider. If (w.l.o.g.) we assume agent 1 to be the first decider, then

$$y_{soc}^{(i)}(T^{(0)}) = \pm\theta_i$$

for all agents $i \neq 1$. In a clique with dichotomous thresholds, we have (supposing a correct first decision) that after the first decision an agent with the lower threshold $\theta_{\min}$ will update their belief by $\theta_{\min}$ and an agent with the higher threshold $\theta_{\max}$ will update their belief by $\theta_{\max}$. To obtain the size of the first wave, we compute what fraction of the γN agents with threshold $\theta_{\min}$ have belief in the interval $[0, \theta_{\min}$ and what fraction of the $(1-\gamma)N$ agents with threshold $\theta_{\max}$ have belief in the interval $[0, \theta_{\max})$. Using these, we have that after a correct first decision

$$\begin{aligned}
\mathbb{E}[a_1|T] &= \gamma N \quad \frac{\int_0^{\theta_{\min}} p_+^*(x,T)dx}{\int_{-\theta_{\min}}^{\theta_{min}} p_+^*(x,T)dx} + (1-\gamma)N \frac{\int_0^{\theta_{\max}} p_+^*(x,T)dx}{\int_{-\theta_{\max}}^{\theta_{\max}} p_+^*(x,T)dx} \\
&\approx \frac{\gamma N}{2} \quad \left(1 + \frac{\sqrt{T/\pi}}{1 - \frac{2}{\theta_{\min}}\sqrt{T/\pi} e^{-\frac{\theta_{\min}^2}{4T}}}\right) \\
&\qquad + \frac{(1-\gamma)N}{2}\left(1 + \frac{\sqrt{T/\pi}}{1 - \frac{2}{\theta_{\max}}\sqrt{T/\pi} e^{-\frac{\theta_{\max}^2}{4T}}}\right)
\end{aligned}$$

where we used the same approximations as in Section 3.4. Substituting $\mathbb{E}[T] \approx \frac{\theta_{\min}^2}{4\ln(\gamma N)}$ for T gives

$$\begin{aligned}
\mathbb{E}[a_1] \approx \frac{\gamma N}{2}&\left(1 + \frac{\frac{\theta_{\min}}{2}}{\sqrt{\pi \ln \gamma N} - \frac{1}{\gamma N}}\right) \\
&+ \frac{(1-\gamma)N}{2}\left(1 + \frac{1}{\frac{2}{\theta_{\min}}\sqrt{\pi \ln \gamma N} - \frac{2}{\theta_{\max}}(\gamma N)^{-(\theta_{\max}/\theta_{\min})^2}}\right).
\end{aligned}$$

For sufficiently large N, this becomes

$$\mathbb{E}[a_1] \approx \frac{N-1}{2}\left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right). \tag{4.3}$$

Figure 4.2 compares Eq. (4.3) with simulation results. By a similar argument, when the first decision is incorrect we have

$$\mathbb{E}[a_1^-] \approx \frac{N-1}{2}\left(1 - \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right).$$

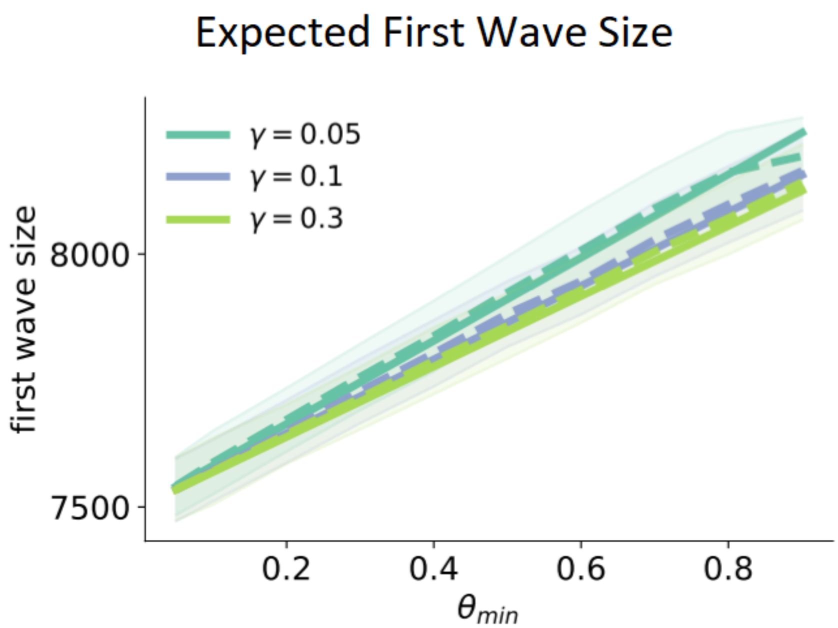

Figure 4.2: First wave size for dichotomously distributed thresholds for various fractions γ of the clique with lower threshold in the consensus bias case. Upper threshold is held constant at $\theta_{\max} = 1$; clique size is held constant at $N = 15000$. Simulations (dashed) and theory (solid) from Eq. (4.3) are shown. Figure taken from [59].

4.3 Expected size of second wave under consensus bias

To determine the size of the second wave, we again need to consider both high and low threshold observers. Let $R_{\pm,k}(t)$ be the information communicated at the end of the first wave cycle by an agent with threshold θ_k. Then we have that

$$\begin{aligned} c_1^+ &= a_1 R_{+,k}(t) + \Big(N - a_1 - 2\Big) R_{-,k} \\ &= \Big(2a_1 - (N-2)\Big) R_{+,k}(t). \end{aligned}$$

Assuming the first decision is correct, we can take a conditional expectation. Substituting the expected value of a_1 given in Eq. (4.3) and the expected value of $R_{+,k}$ given in Eq. (3.8) gives

$$\hat{c}_1^+ \approx \left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right) R_{+,k}(t) \approx \frac{N\theta_{\min}^2}{2\pi \ln(\gamma N)}. \tag{4.4}$$

Similarly, when the first decision is wrong, the expected increment after the first wave is given by

$$\begin{aligned} \hat{c}_1^- &\approx \left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right) R_{+,k}(t) \\ &\approx \frac{N\theta_{\min}^2}{2\pi \ln(\gamma N)}. \end{aligned}$$

In the last panel of Figure 4.1 we compare Eq. (4.4) with simulation results. Interestingly, the equation holds regardless of whether it was an agent with threshold $\theta_{\max}$ or threshold $\theta_{\min}$ who observed the decision.

To explain why, first note that by assuming t is small and using a Taylor expansion for $\ln(1-x)$

around $x = 1$ we can show that

$$R_{+,k}(t) \approx 2\sqrt{\frac{t}{\pi}}\left(e^{-\frac{\theta_k^2}{4t}} + 1\right).$$

Thus, the threshold of the observer only factors into the equation via an exponentially small term. When N is large, that term vanishes and the value of $R_{+,k}(t)$ is dominated by the decision time t, regardless of whether the observing agent believes the threshold they observe is high or low. When the group is large, first decisions occur quickly before the belief distributions can interact with the boundaries. Therefore $\hat{c}_1^{\pm}$ is approximately independent of the observer's threshold.

4.4 Expected accuracy of a random agent in a dichotomous clique under consensus bias

As in homogeneous networks, in our dichotomous networks with consensus bias $\hat{c}_1^{\pm}$ grows with N and when $\hat{c}_1^{\pm} \geq 2\theta_{\max}$, we expect all agents to decide by the end of the second wave. Again following the pattern of homogeneous networks, if the first decision is correct we expect the entire clique to follow (middle panel of Figure 4.1). In the case of a wrong first decision, on average, slightly less than half the clique will follow the wrong decision in the first wave while the rest of the clique will self-correct and make the correct decision in the second wave.

As in Section 3.8, we abuse notation and take $P^{+}(y_i(T^W) \geq \theta_i)$ to be the probability that agent i's belief exceeds the positive threshold during some wave $0 \leq W$ of social updating. This is the probability that agent i makes a correct decision before the end of social updating following the first decision and can again be given as the fraction of the clique deciding correctly by time the waves come to an end.

Recalling our assumptions that the first decider will have the lower threshold $\theta_{\min}$ and that, without loss of generality, the first decider is agent 1, our expected accuracy for a random agent in a sufficiently large clique is given by

$$
\begin{aligned}
\mathbb{E}[P^+\Big(y_i(T^W)\geq\theta_i\Big)] &= P^+(y_1(T)=\theta_1)P^+\Big(y_i(T^W)\geq\theta_i|y_1(T)=\theta_1\Big)\\
&\quad+P^+(y_1(T)=-\theta_1)P^+\Big(y_i(T^W)\geq\theta_i|y_1(T)=-\theta_1\Big)\\
&=\frac{1}{1+e^{-\theta_{\min}}}(1)+\frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}\frac{N-1-\mathbb{E}[a_1^-]}{N}\\
&\approx\frac{1}{1+e^{-\theta_{\min}}}+\frac{e^{-\theta_{min}}}{1+e^{-\theta_{\min}}}\frac{N-1-\frac{N-1}{2}\left(1-\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)}{N}
\end{aligned}
$$

where the approximation in the last line comes from the fact that the formula given for $\mathbb{E}[a_1^-]$ is an approximation based on the assumption that N is large. Using this assumption we can simplify further to obtain

$$
\begin{aligned}
\mathbb{E}[P^+\Big(y_i(T^W)\geq\theta_i\Big)] &\approx\frac{1}{1+e^{-\theta_{\min}}}\\
&\quad+\frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}\left(\frac{N-1}{N}-\frac{(N-1)}{2N}\left(1-\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)\right)\\
&=\frac{1}{1+e^{-\theta_{\min}}}+\frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}\left(\frac{1}{2}+\frac{1}{2}\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right).
\end{aligned}
\tag{4.5}
$$

This expectation is compared to simulation results in the middle panel of Figure 4.1. As $\theta_{\min}$ increases, the assumption that the first decider has the lower threshold $\theta_{\min}$ slightly breaks down, leading to an underperformance of the approximation when compared to simulations. In trials where the first decider has threshold $\theta_{\max}$, an accurate first decision is more likely. Overall, dichotomous cliques under consensus bias behave like homogeneous cliques with threshold $\theta_{\min}$: uninformed agents govern decisions leading to fast, inaccurate choices.

4.5 First wave size in omniscient dichotomous cliques

In contrast, agents in omniscient cliques know other agents' thresholds precisely and update their social evidence according to Eq. (2.10) with boundaries of integration determined using the method described in Section 2.5.

In this case we again have that that $P_{\min} \gg P_{\max}$ for most parameter values, hence, in obtaining the first wave size we again only consider the case where a first decision is made by an agent with threshold $\theta_{\min}$ and use a method similar to that in Section 3.4. Accordingly, we expect about half of the lower threshold group to decide in the first wave.

Assume the first decision is made by an agent with threshold $\theta_{\min}$ and that it is correct. Let $\Delta\theta = \theta_{\max} - \theta_{\min}$. In this case,

$$
\begin{aligned}
\mathbb{E}[a_1|T] &= (\gamma N - 1) \quad \frac{\int_0^{\theta_{\min}} p_+^*(x,T)dx}{\int_{-\theta_{\min}}^{\theta_{\min}} p_+^*(x,T)dx} + (1-\gamma)N \frac{\int_{\Delta\theta}^{\theta_{\max}} p_+^*(x,T)dx}{\int_{-\theta_{\max}}^{\theta_{\max}} p_+^*(x,T)dx} \\
&\approx \frac{(\gamma N - 1)}{2} \quad \left(1 + \frac{\operatorname{erf}\frac{\sqrt{T}}{2}}{\operatorname{erf}\frac{2\theta_{\min}}{2\sqrt{T}}}\right) + \frac{(1-\gamma)N}{2}\left(\frac{\operatorname{erf}\frac{\theta_{\max}}{2\sqrt{T}} - \operatorname{erf}\frac{\Delta\theta}{2\sqrt{T}}}{\operatorname{erf}\frac{\theta_{\max}}{2\sqrt{T}}}\right) \\
&\approx \frac{(\gamma N - 1)}{2} \quad \left(1 + \frac{\frac{\sqrt{T}}{2}}{1 - \frac{2}{\theta_{\max}\sqrt{\frac{T}{\pi}}} e^{-\frac{\theta_{\max}^2}{4T}}}\right) \\
&\qquad\qquad + \frac{(1-\gamma)N}{2}\left(\frac{\frac{2}{\Delta\theta}\sqrt{\frac{T}{\pi}} e^{-\frac{\Delta\theta^2}{4T}} - \frac{2}{\theta_{\max}}\sqrt{\frac{T}{\pi} e^{-\frac{\theta_{\max}^2}{4T}}}}{1 - \frac{2}{\theta_{\max}}\sqrt{\frac{T}{\pi}} e^{-\frac{\theta_{\max}^2}{4T}}}\right).
\end{aligned}
$$

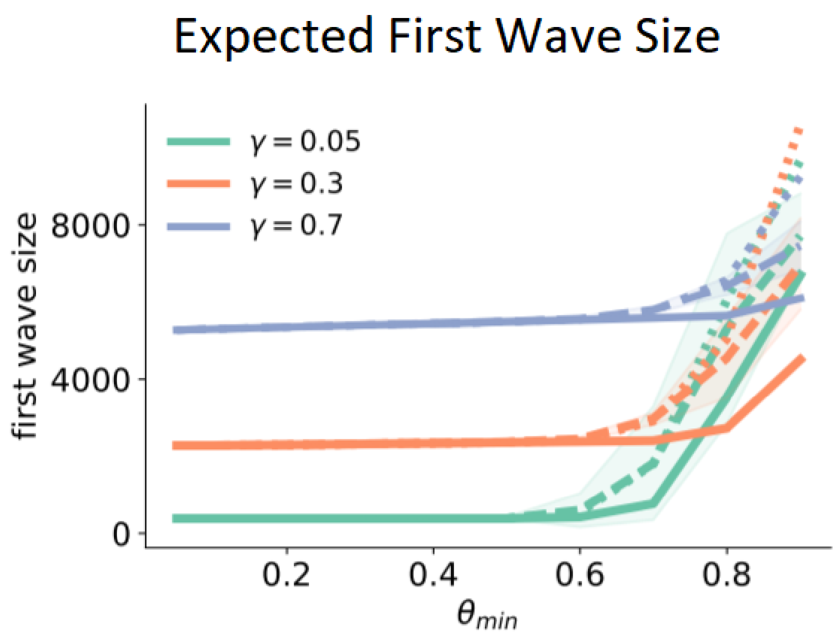

Figure 4.3: First wave size for dichotomously distributed thresholds for various fractions γ of clique at lower threshold in the omniscient case. The upper threshold is held constant at $\theta_{\max} = 1$. Simulations (dashed) and theory (solid) from Eq. (4.6) are shown. Figure from supplemental material for [59].

Substituting in $\mathbb{E}[T] \approx \frac{\theta_{\min}^2}{\sqrt{4\pi \ln(\gamma N)}}$ for T gives

$$\begin{aligned}\mathbb{E}[a_1] &\approx \frac{\gamma N - 1}{2}\left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right) + \\ &\qquad \frac{(1-\gamma)N}{2}\left(\frac{\frac{1}{\Delta\theta}(\gamma N)^{-\left(\frac{\Delta\theta}{\theta_{\min}}\right)^2} - \frac{1}{\theta_{\max}}(\gamma N)^{-\left(\frac{\theta_{\max}}{\theta_{\min}}\right)^2}}{\frac{1}{\theta_{\min}}\sqrt{\pi \ln(\gamma N)} - \frac{1}{\theta_{\max}}(\gamma N)^{-\left(\frac{\theta_{\max}}{\theta_{\min}}\right)^2}}\right) \\ &\approx \frac{\gamma N - 1}{2}\left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right)\end{aligned} \tag{4.6}$$

for large N. Figure 4.3 compares Eq. (4.6) with simulation results.

4.6 Increment after first wave in omniscient dichotomous cliques

Intuitively, we may expect that the evidence revealed by a handful of low-threshold agents may be insufficient to convince agents with higher thresholds. However, if the subpopulation of agents with a lower threshold is sufficiently large, they may cumulatively provide enough evidence due to their decisions or non-decisions in the first wave to trigger decisions of higher threshold agents so that the rest of the clique reaches a decision in the second wave. (See Figure 4.4.) As in the homogeneous case, decisions made in the second wave are expected to be accurate. Ideally, the distribution of dichotomous thresholds could be chosen so that rapid decisions are made by agents with threshold $\theta_{\min}$ but the subset of those agents making incorrect decisions is small enough so that a substantial fraction of the whole clique makes quick and correct decisions.

The goal, then, is to choose a large enough low-threshold subpopulation to guarantee that higher-threshold agents will choose in the second wave without sacrificing too large a percentage of the total (high and low threshold agents) population to following a wrong first decision.

To this end, we seek an expression for the size of the social information increment c_1 to agents after the first wave using methods similar to those used in the previous section.

Assume an agent with threshold $\theta_{\min}$ made the first decision and that it is correct. Then

$$\mathbb{E}[c_1^+|T] = a_{1,\min+}R_{\min,+}(t) + (\gamma N - a_{1,\min+} - j)R_{\min,-}(t)$$

where $j = 1$ if the observing agent has threshold $\theta_{\max}$ and $j = 2$ if the observing agent has

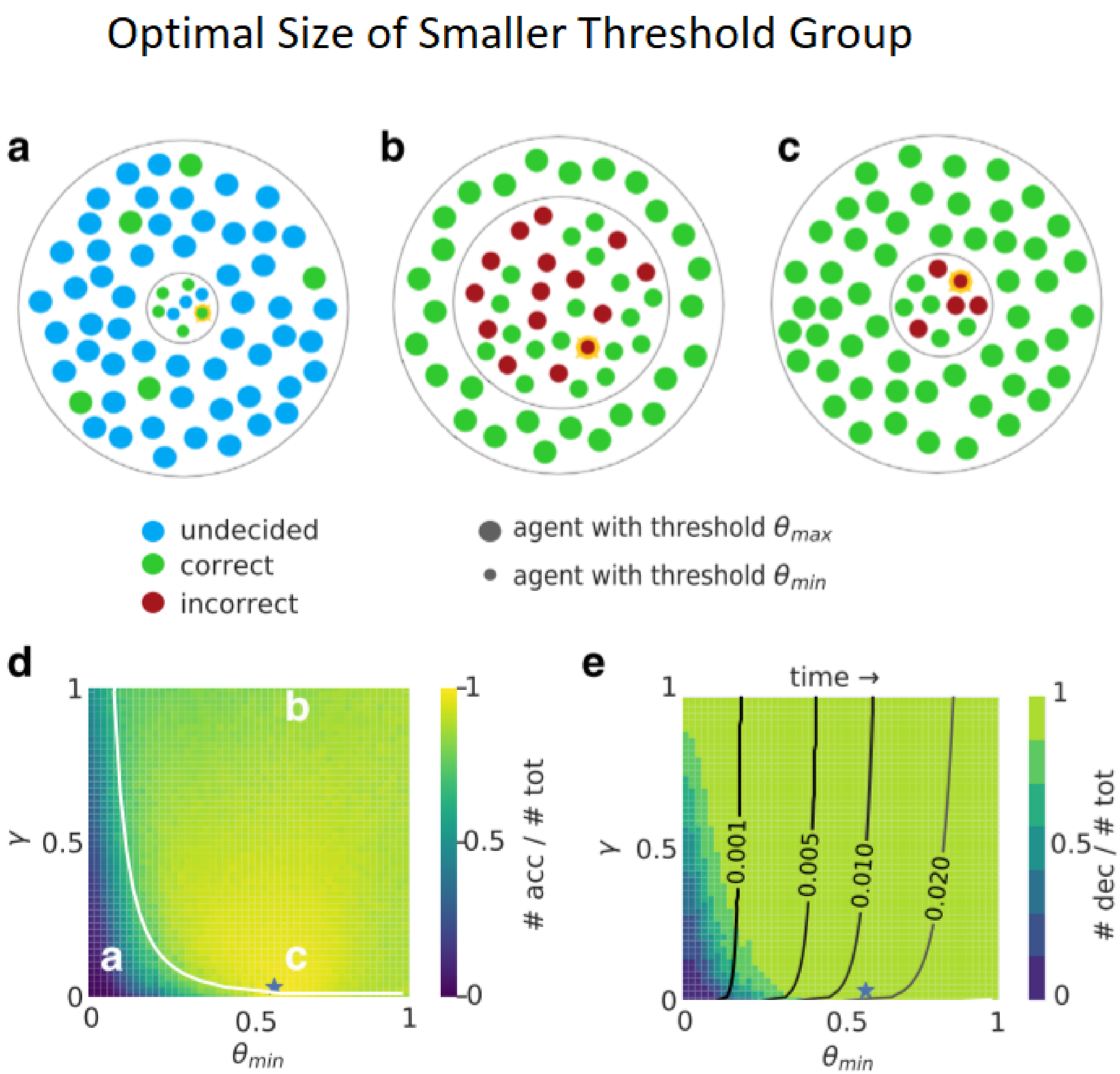

Figure 4.4: (a) With few low-threshold agents, the remaining agents receive insufficient information to decide after the first wave; (b) With many low-threshold agents, a wrong first decision sways much of the network; (c) With the right number of low-threshold agents, a few hasty agents follow an incorrect decision, but the difference between agreeing and disagreeing low-threshold agents drives the rest to choose correctly. (d) Fraction of the clique choosing accurately for a dichotomous threshold clique. White line represents Eq. (4.8). (e) Fraction of the clique deciding by the end of the second wave. Isoclines indicate time to first decision. $N = 15000$ in (b) and (c). Figure taken from [59].

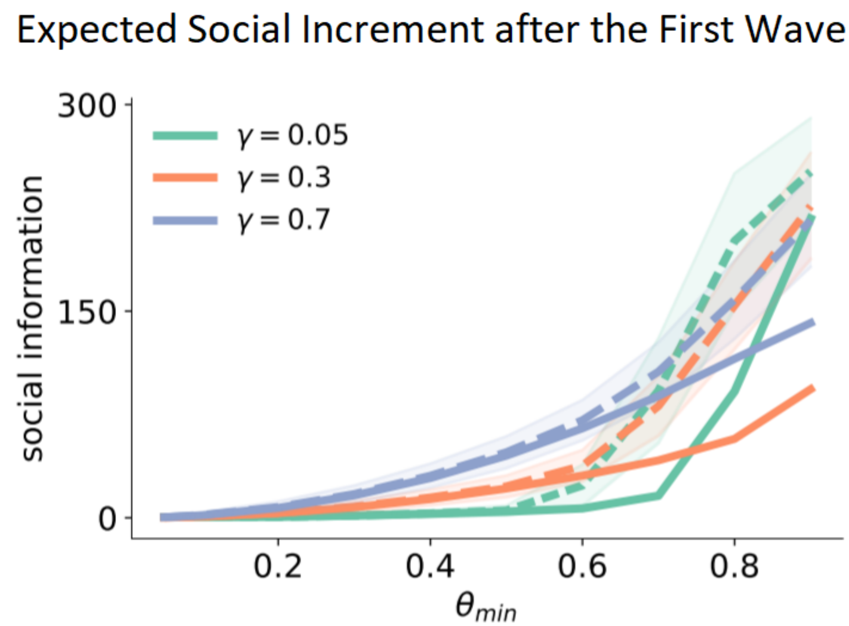

Figure 4.5: Social increment after first wave for dichotomously distributed thresholds for various γ in the omniscient case. The upper threshold is again held constant at $\theta_{\max} = 1$. Simulations (dashed) and theory (solid) from Eq. (4.7) are shown. Simulations and theory match closely while $\Delta\theta = \theta_{\max} - \theta_{\min}$ is large. When this quantity is small (for the greater values of $\theta_{\min}$) the assumption underlying the theoretical approximations (namely, that the first decider has the lower threshold $\theta_{\min}$) breaks down. Figure taken from supplemental material for [59].

threshold $\theta_{\min}$. Hence,

$$\begin{aligned}\mathbb{E}[c_1^+] &\approx \left((\gamma N - 1)\left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right) - (\gamma N - j) \right) \frac{\theta_{\min}}{\sqrt{\pi \ln(\gamma N)}} \\ &\approx \frac{\gamma N \theta_{\min}^2}{2\pi \ln(\gamma N)}.\end{aligned} \tag{4.7}$$

By a similar argument, we can determine that the increment in the case of a wrong first decision is given by the same expression:

$$\mathbb{E}[c_1^-] \approx \frac{\gamma N \theta_{\min}^2}{2\pi \ln(\gamma N)}.$$

Figure 4.5 compares Eq. (4.7) with simulation results.

4.7 Optimal fraction of low threshold agents in omniscient dichotomous cliques

In finite populations we seek values of γ and $\theta_{\min}$ such that the first wave is large enough to convince the remainder of the population, but small enough to buffer the majority of the population from following an incorrect first choice. We thus expect that the population makes the best collective decisions (best balance of speed and accuracy) at intermediate values of γ and $\theta_{\min}$ (star in Figure 4.4d).

If γ is too small, not enough information can be extracted from the behavior of the hasty deciders. If γ is too large, the majority of the clique will choose the correct choice, but too many agents from the hasty group will make the incorrect decision. We thus require $c_1^- = 2\theta_{\max}$. If $c_1^- > 2\theta_{\max}$, then too many agents are sacrificed to the wrong decision and if $c_1^- < 2\theta_{\max}$, not enough information is extracted from hasty deciders to convince the rest of the clique.

Setting $c_1^- = 2\theta_{\max}$,

$$\frac{\theta_{\min}^2 \gamma N}{2\pi \ln(\gamma N)} \approx \frac{\theta_{\min}^2 \gamma N}{2\pi \ln(N)} = 2\theta_{\max}.$$

Solving for γ, we have

$$\gamma \approx \frac{4\pi\theta_{\max}}{N} \frac{\ln N}{\theta_{\min}^2} \tag{4.8}$$

as the optimal fraction of hasty deciders in a clique with dichotomously distributed thresholds of size N. The function in Eq. (4.8) is represented by the white line in Figure 4.4d and predicts the true optimal fraction of low threshold agents well.

In finite populations both γ and $\theta_{\min}$ must be large enough to convince the remainder of the population (Figure 4.4 a), but small enough to buffer the rest of the population from an incorrect first choice (Figure 4.4 b.) The above expression for γ represents a balance between these

Fraction of Agents Deciding Accurately

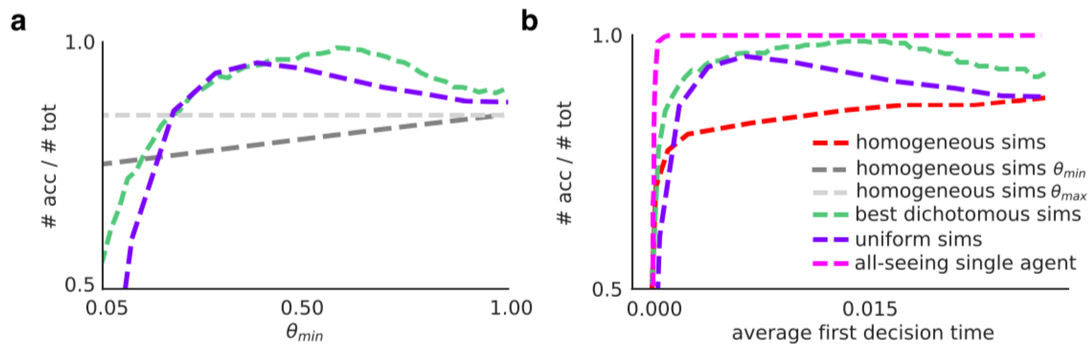

Figure 4.6: (a) Mean fraction of clique choosing accurately after two waves for different threshold distributions in omniscient populations. For the dichotomous case, the fraction γ of the clique at lower threshold is chosen to maximize accuracy for each θ_{min} value. (b) Over a range of possible first decision times, heterogeneous cliques give better accuracy than homogeneous ones with omniscient social updating (See Section 4.12 for simulation details). Also shown: accuracy of a single agent with access to private information of all agents. $N = 15000$ and $\theta_{\text{max}} = 1$. Figure taken from [59].

eventualities that enables nearly all agents to decide by the end of the second wave (Figure 4.4 e) while providing maximal accuracy (star, white line in Figure 4.4 d.) Finite populations with dichotomous thresholds can sacrifice a small fraction of early adopters so the majority makes a fast, correct choice. Agents in heterogenous networks can thus decide more quickly and outperform agents in homogeneous networks in recovering from a wrong first choice (Figure 4.4 d and e; Figure 4.6).

4.8 Expected accuracy of a random agent in an omniscient dichotomous clique

Assume that γ is large enough in relation to θ_{min} that the entirety of the clique has decided by the end of the second wave, and that the first decider (agent 1) has threshold $\theta_1 = \theta_{\text{min}}$. We would like to find $P^+(y_i(T^W) \geq \theta_i)$, which we define as the probability that agent i's belief has crossed the positive threshold during some wave $0 \leq W$. Then (following the method in Section 4.4) we have that the expected fraction of an omniscient clique with dichotomous thresholds that decides accurately

$$\begin{aligned}
\mathbb{E}[P+\Big(y_i(T^W) \geq \theta_i\Big)] &= P^+(y_1(T) = \theta_1)\mathbb{E}[P^+\Big(y_i(T^W) \geq \theta_i | y_1(T) = \theta_1\Big)] \\
&\quad + P^+(y_1(T) = -\theta_1)\mathbb{E}[P+\Big(y_i(T^W) \geq \theta_i | y_1(T) = -\theta_1\Big)] \\
&= \frac{1}{1+e^{-\theta_{\min}}} + \frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}} \frac{N-1-\mathbb{E}[a_1^-]}{N} \\
&\approx \frac{1}{1+e^{-\theta_{\min}}} + \frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}} \frac{N-1-\frac{(\gamma N-1)}{2N}\left(1-\frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right)}{N} \\
&\approx \frac{1}{1+e^{-\theta_{\min}}} + \frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}} \left((1-\frac{\gamma}{2}) + \frac{\gamma}{2} \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right).
\end{aligned} \tag{4.9}$$

Thus, unlike homogeneous groups or heterogeneous groups under consensus bias, the accuracy of the group depends not only on the size of the thresholds but also on the fraction of the clique at the smaller threshold.

4.9 First decision times for uniformly distributed threshold

The general conclusions for populations with two threshold values also hold for populations with more complicated distributions. In the next few sections we show that agents with thresholds selected from a uniform distribution behave under consensus bias similarly to populations with dichotomous threshold distributions under consensus bias. Analysis of omniscient agents is unfortunately more complicated and we will present numerical results.

Assume each agent has a threshold sampled from a uniform distribution on $[\theta_{\min}, \theta_{\max}]$. To find an expectation for the first decision time in such a clique, let $\Phi_j(t)$ be the be the lifetime distribution function for a particle evolving according to Eq. (2.2) on the domain $(-\theta_j, \theta_j)$ and let $\rho_j(t)$ be the corresponding first passage time density so that $\Phi_j(t) = \int_0^t \rho_j(s)ds$. Then the first decision time distribution is given by

$$\begin{aligned}
p_N(t) &= -\frac{d}{dt}\left(\prod_{i=1}^{N}(1-\Phi_i(t))\right) \\
&= \sum_{i=1}^{N}\left(\prod_{j\neq i}(1-\Phi_j(t))\right)\rho_i(t) \\
&= \sum_{i=1}^{N}\exp\left(\log\left(\prod_{j\neq i}(1-\Phi_j(t))\right)\right)\rho_i(t) \\
&= \sum_{i=1}^{N}\exp\left(\sum_{j\neq i}\log\left(1-\Phi_j(t)\right)\right)\rho_i(t) \\
&\approx \sum_{i=1}^{N}\exp\left(-\sum_{j\neq i}\Phi_j(t)\right)\rho_i(t).
\end{aligned}$$

We now replace $\exp\left(-\sum_{j\neq i}\Phi_j(t)\right)$ with $\exp\left(-\sum_k \Phi(t,\theta_k)\right)$ in the last line above where $\theta_k = \theta_{min} + \frac{\Delta\theta}{N}$ for $k \in \{1, 2, ...N\}$. That is, for sufficiently large N, thresholds are distributed evenly in $[\theta_{min}, \theta_{max}]$. Multiplying and dividing $-\sum_{j\neq i}\Phi(t,\theta_k)$ by $\frac{\Delta\theta}{N}$ yields a Riemann sum. As $N \to \infty$, this approaches an integral. Hence,

$$p_N(t) \approx \frac{N}{\Delta\theta}\exp\left(-\frac{N}{\Delta\theta}\int_{\theta_{\min}}^{\theta_{max}}\Phi(t,\theta)d\theta\right)\int_{\theta_{\min}}^{\theta_{\max}}\rho(t,\theta)d\theta.$$

We can compute explicitly that

$$\int_{\theta_{\min}}^{\theta_{\max}} \Phi(t,\theta)d\theta = \frac{1}{2}\Bigg[4\sqrt{\frac{t}{\pi}}\Bigg(e^{-\frac{\theta_{\min}^2+t^2}{4t}}\cosh\big(\frac{\theta_{\min}}{2}\big) - e^{-\frac{\theta_{\max}^2+t^2}{4t}}\cosh\left(\frac{\theta_{\max}}{2}\right)\Big) - (t+\theta_{\min}+e^{\theta_{\min}}+1)\,\mathrm{erfc}\left(\frac{\theta_{\min}+t}{2\sqrt{t}}\big) + (t+-\theta_{\min}+e^{-\theta_{\min}}+1)\,\mathrm{erfc}\left(\frac{\theta_{\min}-t}{2\sqrt{t}}\right) + (t+\theta_{\max}+e^{\theta_{\max}}+1)\,\mathrm{erfc}\left(\frac{\theta_{\max}+t}{2\sqrt{t}}\right) + (\theta_{\max}-e^{-\theta_{\max}}-t-1)\,\mathrm{erfc}\left(\frac{\theta_{\max}-t}{2\sqrt{t}}\right)\Bigg].$$

Computing the derivative with respect to time we obtain

$$\int_{\theta_{\min}}^{\theta_{\max}} \rho(t,\theta)d\theta = \frac{1}{2\sqrt{\pi t}}\Bigg[2e^{-\frac{(\theta_{\min}+1)^2}{4t}}(1+e^{\theta_{\min}}) - 2e^{-\frac{(\theta_{\max}+1)^2}{4t}}(1+e^{\theta_{\max}}) + \sqrt{\pi t}\left(\mathrm{erf}\left(\frac{-\theta_{\min}+t}{2\sqrt{t}}\right)+\mathrm{erf}\left(\frac{\theta_{\min}+t}{2\sqrt{t}}\right)\right) - \mathrm{erf}\left(\frac{-\theta_{\max}+t}{2\sqrt{t}}\right) - \mathrm{erf}\left(\frac{\theta_{\max}+t}{2\sqrt{t}}\right)\Bigg].$$

We can now calculate $\mathbb{E}[T]$ for the uniform distribution. In this case,

$$\begin{aligned} P(\tau_N^* > t) &\approx \prod_{i=1}^{N}\mathrm{erf}\left(\frac{\theta_i}{2\sqrt{t}}\right) \\ &= \exp\left(\sum_{i=1}^{N}\log\left(\mathrm{erf}\left(\frac{\theta_i}{2\sqrt{t}}\right)\right)\right) \\ &\approx \exp\left(-\,\mathrm{erfc}\left(\frac{\theta_i}{2\sqrt{t}}\right)\right) \end{aligned}$$

where in the last line we have used that, for small t, $\mathrm{erf}\left(\frac{\theta_i}{2\sqrt{t}}\right) \approx 1$. The sum in the above equation can be replaced with integrals with the appropriate scaling. Hence,

$$
\begin{aligned}
P(\tau_N^* > t) &\approx \exp\left(-\frac{N}{\Delta\theta} \int_{\theta_{\min}}^{\theta_{\max}} \operatorname{erfc}\left(\frac{\theta}{2\sqrt{t}}\right) d\theta\right) \\
&= \exp\left(-\frac{N}{\Delta\theta} \left(2\sqrt{\frac{t}{\pi}}\left(e^{-\frac{\theta_{\min}^2}{4t}} - e^{-\frac{\theta_{\max}^2}{4t}}\right)\right.\right. \\
&\qquad \left.\left. + \theta_{max} \operatorname{erfc}\left(\frac{\theta_{\max}}{2\sqrt{t}}\right) - \theta_{\min} \operatorname{erfc}\left(\frac{\theta_{\min}}{2\sqrt{t}}\right)\right)\right) \\
&\approx \exp\left(-\frac{4t^{3/2}}{\sqrt{\pi}} \frac{N}{\Delta\theta}\left(\frac{e^{-\frac{\theta_{\min}^2}{4t}}}{\theta_{\min}^2} - \frac{e^{-\frac{\theta_{\max}^2}{4t}}}{\theta_{\max}^2}\right)\right) \\
&\approx \exp\left(-\frac{4t^{3/2}}{\sqrt{\pi}} \frac{N}{\Delta\theta}\left(\frac{e^{-\frac{\theta_{\min}^2}{4t}}}{\theta_{\min}^2}\right)\right)
\end{aligned}
$$

where in the last line we have assumed that the larger term dominates when t is very small.

We now define a sequence t_N such that $P(\tau_N^* > t_N) = p$ for some $0 < p < 1$. In this case,

$$t_N = \frac{\theta_{\min}^2}{6W(\Lambda)}$$

where $W(x)$ is the Lambert-W function and

$$\Lambda^3 = \frac{2\theta_{min}^5 N^2}{27(\ln p)^2 \Delta\theta^2 \pi}.$$

Since $W(x) \approx \ln x$ for large x, we have (after some algebra)

$$\frac{1}{\pi \ln(p)^2} = \frac{27 \exp\left(\frac{\theta_{\min}^2}{2t_N}\right) \Delta\theta^2}{2\theta_{\min}^5 N^2}$$

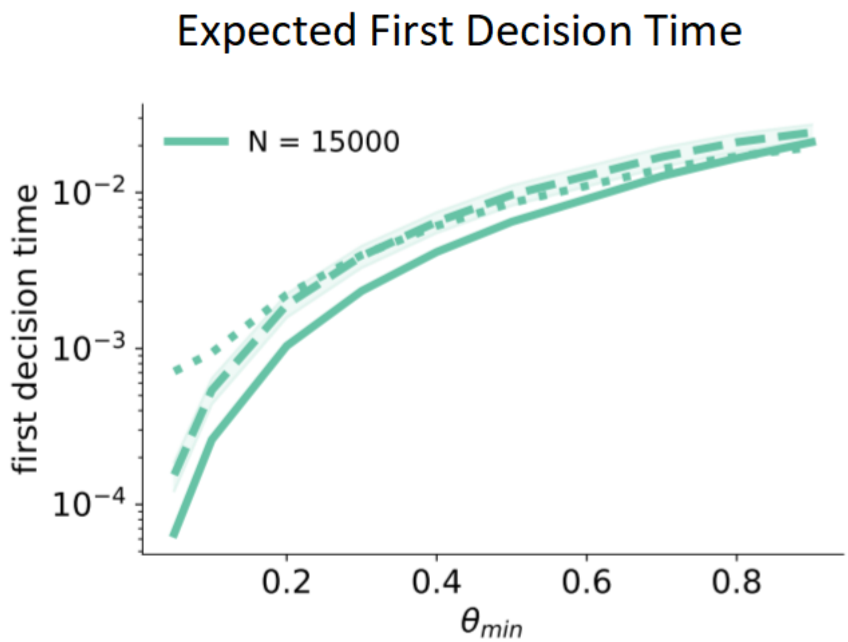

Figure 4.7: Mean first passage time in the uniformly distributed thresholds case. Simulations (dashed) and progressive approximations of the mean first passage time, Eq. (4.10). Center approximation (dotted) and right approximation (solid) of Eq. (4.10) are shown. Here, $\theta_{\max} = 1$. Figure taken from supplemental material for [59].

so that

$$\begin{aligned}\mathbb{E}[\tau] &= \frac{\theta_{\min}^2}{2\ln\left(\frac{2\theta_{\min}^5 N^2}{27\Delta\theta^2}\right)} \\ &\approx \frac{\theta_{\min}^2}{4\ln N}.\end{aligned} \tag{4.10}$$

Figure 4.7 compares Eq. (4.10) with simulation results.

4.10 First wave size for uniform threshold cliques under consensus bias

Suppose agent i makes the first decision. Then the size of the first wave is

$$\mathbb{E}[a_1|T,\theta_i] = \sum_{j=1,j\neq i}^{N} \frac{\int_0^{\theta_j} p_+^*(x,T)dx}{\int_{-\theta_j}^{\theta_j} p_+^*(x,T)dx}.$$

Proceeding according to methods in previous sections,

$$\begin{aligned}\mathbb{E}[a_1|\theta_i] &\approx \sum_{j=1,j\neq i}^{N} \frac{1}{2}\left(1+\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)\\ &= \frac{N-1}{2}\left(1+\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right).\end{aligned}$$

Finally,

$$\begin{aligned}\mathbb{E}[a_1] &\approx \frac{N-1}{2}\left(1+\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)\sum_{i=1}^{N} p_{i,+}\\ &= \frac{N-1}{2}\left(1+\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)\end{aligned} \tag{4.11}$$

where $p_{k,+}$ is the probability that the first decider has threshold θ_k given that the first decision is correct; this probability can be computed as in Section 4.1. By a similar argument, we can show that when the first decision is incorrect,

$$\begin{aligned}\mathbb{E}[a_1] &\approx \frac{N-1}{2}\left(1-\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)\sum_{i=1}^{N} p_{i,-}\\ &= \frac{N-1}{2}\left(1-\frac{\theta_{\min}}{\sqrt{4\pi\ln(\gamma N)}}\right)\end{aligned} \tag{4.12}$$

Thus, a clique with thresholds drawn from a uniform distribution under consensus bias, like a clique with dichotomous thresholds under consensus bias, has a first wave whose size is consistent with the first wave size of a homogeneous clique with threshold $\theta = \theta_{\min}$.

4.11 Social increment following first wave for uniform threshold cliques under consensus bias

Since the clique is operating under consensus bias, an observer with threshold θ_i will assume all other agents also have threshold θ_i. Hence, the amount of social information an agent receives depends on who is observing. In this case, after a correct first decision at time T the expected social increment received by agent i will be

$$\mathbb{E}[c_{1,i}^{+}|T] = \mathbb{E}[a_1]R_{i,+}(T) + (N - \mathbb{E}[a_1] - 2)R_{i,-}.$$

Proceeding as we did in previous sections, we have

$$\begin{aligned} \hat{c}_1^{+} &\approx (N-1)\left(1 + \frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right)\frac{\theta_{\min}}{\sqrt{\pi \ln(N)}} - (N-2)\frac{\theta_{\min}}{\sqrt{\pi \ln(N)}} \\ &\approx \frac{N\theta_{\min}^2}{2\pi \ln(N)}. \end{aligned} \tag{4.13}$$

As this final quantity does not depend on i, all undecided agents' updates in the second wave are given by this equation. By a similar argument, we can show that $\hat{c}_1^{-}$ also has the same amount. Figure 4.8 compares Eq. (4.13) with simulation results.

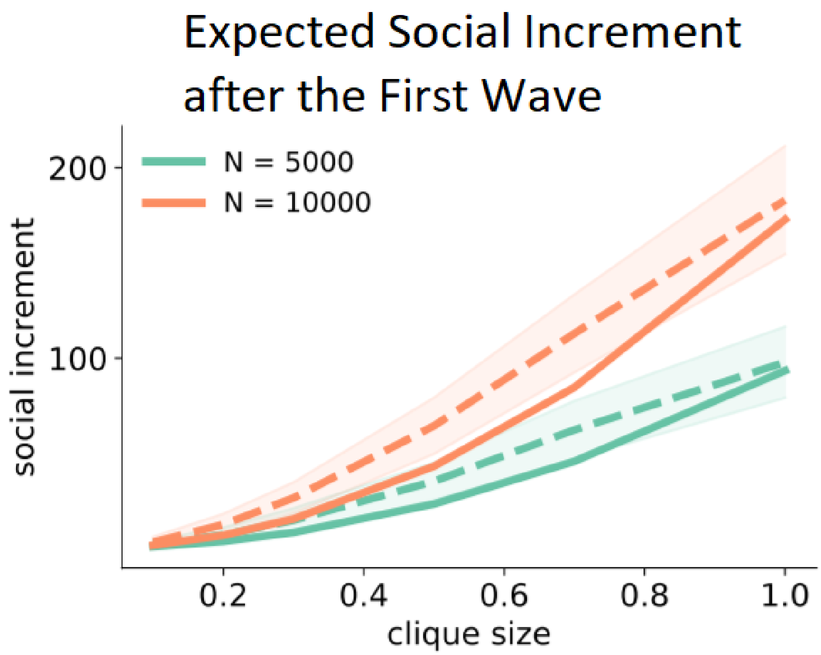

Figure 4.8: Increment in second wave for uniformly distributed thresholds in the consensus bias case. Simulations (dashed) and theory (solid) from Eq. (4.13) from are shown. Here, $\theta_{\min} = 0.3$ and $\theta_{\max} = 1$. Figure taken from supplemental material for [59].

4.12 Conclusion: improved performance for heterogeneous cliques

Previously we found expressions for the fractions of cliques deciding accurately for homogeneous, dichotomous consensus bias, and dichotomous omniscient groups (Eqs. (3.13), (4.5), (4.9) respectively.)

As $N \to \infty$ we have that for both homogeneous and dichotomous consensus bias cliques (noting that in homogeneous cliques $\theta_i = \theta_{\min} = \theta$),

$$\lim_{N\to\infty} \mathbb{E}[P^+\left(y_i(T^W) \geq \theta_i\right)] = \frac{1}{1+e^{-\theta_{\min}}} + \frac{1}{2}\left(\frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}\right).$$

For comparison, assume a dichotomous clique with optimal γ for a given $\theta_{\min}$ value. Then

$$\begin{aligned}\mathbb{E}[P^+\Big(y_i(T^W) \geq \theta_i\Big)] &\approx \frac{1}{1+e^{-\theta_{\min}}} + \frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}\left((1 \quad - \frac{\gamma}{2}) + \frac{\gamma}{2}\frac{\theta_{\min}}{\sqrt{4\pi \ln(\gamma N)}}\right)\\ &\approx \frac{1}{1+e^{-\theta_{\min}}} + \frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}\left(1 \quad - \frac{2\pi\theta_{\max}\ln(N)}{\theta_{\min}^2 N}\right.\\ &\qquad \left. + \frac{2\pi}{\theta_{\min}\sqrt{4\pi\ln\left(\frac{4\pi\theta_{\max}\ln(N)}{\theta_{\min}^2}\right)}}\right)\end{aligned} \tag{4.14}$$

so that

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}[P^+\Big(y_i(T^W) \geq \theta_i\Big)] &= \frac{1}{1+e^{-\theta_{\min}}} + \frac{e^{-\theta_{\min}}}{1+e^{-\theta_{\min}}}(1)\\ &= 1.\end{aligned}$$

Figure 4.6 gives simulation results comparing accuracy and decision time for homogeneous cliques and omniscient cliques with heterogeneous thresholds. To obtain panel b, we fix a time T and determine the threshold θ^* that produces mean first decision time T. The homogeneous plot shows fraction of accurate deciders for that threshold. The uniform curve is obtained by determining what value of $\theta_{\min}$ gives time T. The agents' thresholds are then uniformly distributed on $[\theta_{\min}, \theta_{\max}]$. The dichotomous curve is obtained by determining what values of $\theta_{\min}$ and γ produce the given time T and selecting the values that give highest clique accuracy. The plot shows best performance for thresholds distributed on $[\theta_{\min}, \theta_{\max}]$ uniformly or dichotomously.

Chapter 5

Introducing correlated noise

Our work in the last two chapters showed that cliques can correct for the social evidence kick generated by a wrong first decision, enabling some majority of the clique to decide correctly even in the face of initial social evidence supporting an inaccurate decision. However, it relied on a very strong model assumption that all agents received independent private information. In many situations this assumption is less than realistic: animals choosing where to forage may observe many of the same environmental signals such as smells and sights. Humans choosing whether to adopt a product may read the same product reviews. Due to story-sharing institutions such as the Associated Press, humans who believe they are using individually distinct news sources may still encounter the same story.

In the following chapters, we explore a discrete model in which some of the information agents receive is individual and some common, but agents are unaware of which information is which. Unlike list references, we do not focus on the effect of any form of social information. Instead, we investigate the effect a mix of independent and common private information may have on the accuracy of individual deciders.

5.1 Discrete correlated-evidence case model

We introduce a discrete model similar to the one described in Section 2.1 that also incorporates correlated evidence. In the independent-evidence case model from Section 2.1, at each timestep, i each agent makes a private observation, η_i, of the environment. The posterior probability, *i.e.*, the belief of each agent about which of the two choices is correct, is then computed based on this private evidence. In contrast to previous chapters, we assume that the observers are *not* interacting, and cannot observe each others decisions.

In natural environments, agents often have access to overlapping sources of information. It is therefore unlikely that observations are purely independent. To capture such correlations in the evidence we make a simple change in our assumptions about the statistical structure of the observations: In a network of N agents at each timestep, i, either all agents make the same observation, or all agents make independent observations. In other words, there are $N+1$ sources of information, N of which are each accessible only to one separate agent (the private source), and one source which is accessible to all agents (the common source). On each timestep, all agents sample one observation from one evidence source: With probability $1-c$, all agents sample from their independent private sources, and with probability c they all make the same observation from the common source. Correlated observations and independent observations are drawn from the same observation set, $\{\eta\}$. This could be extended in many ways: Agents could choose the two sources independently of each other, or different subsets of agents could have access to different pools of common information. The analysis in such cases becomes more complex, but the main results about the accuracy of the agents we present below remain similar.

As in the independent-evidence case model, observations in the correlated-evidence case model are drawn until one or more agents reach(es) one of a pair of symmetric thresholds $\{-\theta, \theta\}$. If a single agent reaches threshold first, we call them the 'first decider.' If multiple agents reach the threshold simultaneously, the 'first decider' is chosen with equal probability from that set. The probability of a correct first decision then equals the probability that this 'first decider' makes the correct choice, *i.e.*, that the belief of the first decider reaches the threshold, $\pm\theta$, whose sign agrees with that of the environment $H^{\pm}$ environment. Equivalently, we are computing the probability that on a trial a randomly chosen first decider makes a correct choice.

We remark that there are different ways of defining the probability that the first decision is correct. We call the method outlined above Option a. An alternative possibility, Option b, is to ask about the probability of a correct choice by a first decider chosen with equal probability from all first deciders pooled across different trials. The result is different from the probability of a correct decision by an agent chosen at random from all first deciders on a single trial.

When applying this model, we found an unexpected but interesting phenomenon. To understand this well, note that when fixing the thresholds $\pm\theta$, the probability a single agent makes a correct choice is simply set by their log-likelihood ratio at decision time, so that

$$LLR = \log \frac{P(\text{choice correct})}{1 - P(\text{choice correct})} = \theta \quad \Rightarrow \quad P(\text{choice correct}) = \frac{1}{1 + \mathrm{e}^{-\theta}}. \tag{5.1}$$

On the other hand, if multiple agents randomly accumulate evidence from both private and common pools, the first decider will have accuracy less than the above bound. Understanding and quantifying the origin of this accuracy decrease as well as the consequences for later deciders' accuracy is the primary focus of this chapter and the next.

Figure 5.1 gives the probability (obtained via Monte Carlo simulations) of an accurate first decision as a function of observers choosing a common observation, c, for the options outlined above. The graph on the left assumes highly informative observations so that there is a high probability that the set of agents at threshold at the time of the first decision will contain more than one agent, while the graph on the right assumes minimally informative observations so that the

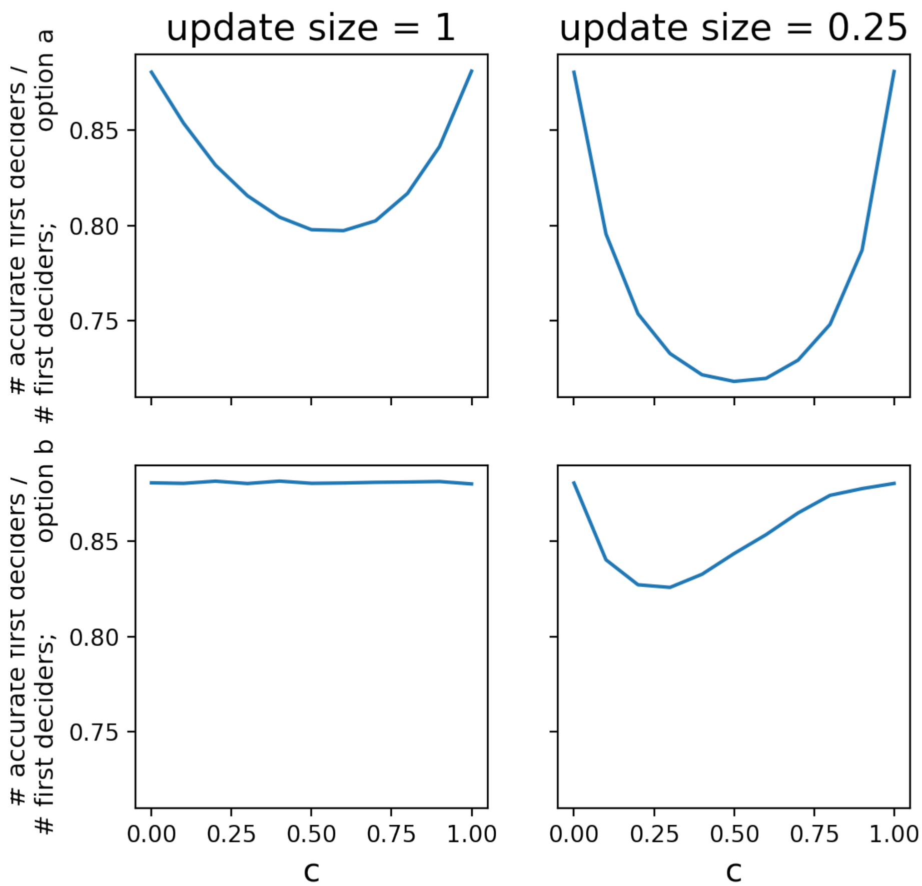

Figure 5.1: Comparison of methods of choosing first decider when more than one agent reaches threshold simultaneously. Option a chooses one agent at random from the pool of deciders on each trial to be counted as the 'first decider'. Option b counts all simultaneous deciders from each trial as first deciders. When this group is large, option b gives the expected accuracy for a single decider $1/(1+e^{-\theta})$. As the update size shrinks (left to right), the expected size of this pool of simultaneous deciders also shrinks and option b begins to exhibit the dip in accuracy characteristic of option a. We note that even for option a accuracy decreases as update size decreases; see Section 6.8 for an explanation of the mechanism behind this phenomenon. Figures are for group size $N = 100$ and threshold $\theta = 2$.

probability of only one agent reaching threshold at the time of the first decision is high. Note how the behavior of the second method shifts.

As in the independent-evidence case, the information obtained from a single observation is a function of the log likelihoods that the observation was sampled from a distribution of measurements conditioned on the state of the environment, H^+ or an H^-. For simplicity, we restrict the set of possible observations from which both correlated and independent observations are drawn from the set $\{\eta_-, \eta_+\}$, where $\eta_\pm$ provides evidence in favor of an $H^\pm$ environment. We let $P(\eta^\pm|H^\pm) = p$ and $P(\eta^\pm|H^\mp) = q$ with $p+q=1$ and $q<p$. Thus, p is a probability of making an observation consistent with the environmental state, while q is the probability of an inconsistent observation.

The sizes of LLR (belief) increments, that is the amount of information provided by a single observation, are thus given by

$$\mathrm{LLR}(\eta^\pm) = \log \frac{P(\eta^\pm \mid H^+)}{P(\eta^\pm \mid H^-)}. \tag{5.2}$$

More explicitly, we can show that the LLR increments are equal in amplitude and opposite in sign (symmetric),

$$\begin{aligned} \mathrm{LLR}(\eta^+) &= \log \frac{P(\eta^+ \mid H^+)}{P(\eta_+ \mid H^-)} = \log \frac{p}{q}; \\ \mathrm{LLR}(\eta^-) &= \log \frac{P(\eta^- \mid H^+)}{P(\eta_- \mid H^-)} = \log \frac{q}{p} = -\log \frac{p}{q} = -\mathrm{LLR}(\eta^+), \end{aligned} \tag{5.3}$$

so that the amount of information given by a single observation is dependent on the relative sizes of p and q.

In the left column of Figure 5.1 we assumed highly informative observations. Assuming $q = p/e$ (as we do throughout Chapter 6), gives $p + p/e = 1$ so that $p = e/(1+e)$, $q = 1/(1+e)$, and hence, $\mathrm{LLR}(\eta^{\pm}) = \pm 1$. In particular, when $q = p/a$ then the information provided by observation $\eta_{\pm}$ equals $\pm \ln a$. As a result, the belief of each agent, y_i, lies on a lattice defined by $\{n \ln a\}_{n \in \mathbb{Z}}$, and we can use the mapping $n \to n \ln a$ or a logarithm in base a to place beliefs on an integer lattice. In the limit of infinitesimal $\ln a$, that is in the limit $a \to 1^+$, we recover a continuous model as outlined in the next section.

5.2 Deriving the continuous correlated-evidence case model

More generally, we can let $f_+(\xi)$ be a probability distribution of observations, ξ, over an arbitrary set Ξ obtained in state H^+, and $f_-(\xi)$ the probability distribution of observations over that same set Ξ in state H^-. Note if the sets of observations for either state differ, then there will be infinitely informative observations which would subverts a typical evidence accumulation process. As previously, we use y to denote accumulated LLR so that in the discrete case we have

$$y(t) = \sum_{s \leq t} \mathrm{LLR}(\xi_s), \tag{5.4}$$

where ξ_s is the observation obtained at time $s \leq t$. Similarly, in the continuous case

$$y(t) = \int_0^t \frac{dy(s)}{ds} ds \tag{5.5}$$

where $\frac{dy}{ds}$ is given by the stochastic drift-diffusion equation described in the previous chapters.

We again assume that in a group of N observers each observer at each timestep t makes an independent private observation with probability $1 - c$, and all observers make a common observation with probability c. Private and common observations have the same conditional distributions, $f_{\pm}(\xi)$ given the state $H^{\pm}$. In the discrete correlated-evidence case, the increment in the belief of the i^{th} observer, $\Delta y_i = y_{i,t+1} - y_{i,t}$, is thus given by the LLR acting on an observation $\xi_{i,t}$

$$\begin{aligned}\Delta y_{i,t} &= \log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \\ &= (1-c)\ \log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} + c\ \log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})},\end{aligned} \tag{5.6}$$

where we are anticipating the fraction of time this increment is drawn from an independent $(1-c)$ or correlated (c) source. Expanding to include conditioning on the state of the environment H,

$$\begin{aligned}\Delta y_{i,t} = (1-c)&\left(\mathbb{E}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right] + \log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} - \mathbb{E}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right]\right) \\ &+ c\left(\mathbb{E}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right] + \log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} - \mathbb{E}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right]\right).\end{aligned} \tag{5.7}$$

We can thus approximate the update in the limit of rapid and infinitesimally weak observations using the Functional Central Limit Theorem

$$\Delta y_{i,t} \approx h_{\Delta t}(t)\Delta t + \sqrt{\Delta t}(\rho_{1-c,\Delta t}(t)\eta_{1-c} + \rho_{c,\Delta t}(t)\eta_c)$$

where η_c and η_{1-c} are random variables with standard normal distributions, and

$$\begin{aligned}h_{\Delta t}(t) &= \frac{1}{\Delta t}\mathbb{E}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right]; \\ \rho^2_{1-c,\Delta t}(t) &= \frac{(1-c)}{\Delta t}\mathrm{Var}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right]; \\ \rho^2_{c,\Delta t}(t) &= \frac{c}{\Delta t}\mathrm{Var}_\xi\left[\log \frac{f_+(\xi_{i,t})}{f_-(\xi_{i,t})} \mid H^\pm\right].\end{aligned} \tag{5.8}$$

Clearly, the drift $h_{\Delta t}$ and the variances $\rho^2_{c,\Delta t}$, $\rho^2_{1-c,\Delta t}$ will diverge unless $f_\pm(\xi)$ are properly scaled in the $\Delta t \to 0$ limit.

We choose a specific scaling for the drift and variances arising from each observation, $\xi_{i,t}$, to ensure the limit holds. Suppose that over a time interval of duration Δt, an observation $\xi_{i,t}$ is the result of $\mu \Delta t$ separate observations. We define a family of stochastic processes parameterized by k, the number of subintervals into which we divide the time increment Δt. Assuming μ is large and $k > 1$, each of the k subintervals contains roughly $\mu_k \equiv \lfloor \mu \Delta t / k \rfloor$ observations with mean and variance that scale linearly with $\mu_k \propto \Delta t / k$. We can achieve this by approximating $\log \frac{f_+(\xi_t)}{f_-(\xi_t)}$ in Eq. (5.7) with a family of stochastic processes parameterized by k:

$$\Delta y_t = \sum_{l=1}^{k} \frac{\Delta t}{k} \log \frac{f_+(\xi_l)}{f_-(\xi_l)} + \frac{\sqrt{c\Delta t}}{\sqrt{k}} \left(\log \frac{f_+(\xi_l)}{f_-(\xi_l)} - \mathbb{E}_\xi \left[\log \frac{f_+(\xi_l)}{f_-(\xi_l)} \mid H^\pm \right] \right) \Bigg)$$

$$+ \frac{\sqrt{(1-c)\Delta t}}{\sqrt{k}} \left(\log \frac{f_+(\xi_l)}{f_-(\xi_l)} - \mathbb{E}_\xi \left[\log \frac{f_+(\xi_l)}{f_-(\xi_l)} \mid H^\pm \right] \right) \Bigg)$$

By the central limit theorem, as $k \to \infty$, the above converges in distribution to

$$\Delta y_t \approx \Delta t h_{\Delta t}(t) + \sqrt{\Delta t}(\rho_{1-c,\Delta t}(t)\eta_{1-c} + \rho_{c,\Delta t}(t)\eta_c)$$

Taking $\Delta t \to 0$ gives

$$dy = hdt + \rho_{1-c} dW + \rho_c dW_c \tag{5.9}$$

where

$$
\begin{aligned}
h(t) &= \lim_{\Delta t \to 0} h_{\Delta t}(t) = \mathbb{E}_{\xi}\left[\log \frac{f_+(\xi_n)}{f_-(\xi_n)}\Big| H^{\pm}\right] \\
\rho_c^2(t) &= \lim_{\Delta t \to 0} \rho_{c,\Delta t}^2(t) = c\mathrm{Var}_{\xi}\left[\log \frac{f_+(\xi_n)}{f_-(\xi_n)}\Big| H^{\pm}\right] \\
\rho_{1-c}^2(t) &= \lim_{\Delta t \to 0} \rho_{c,\Delta t}^2(t) = (1-c)\mathrm{Var}_{\xi}\left[\log \frac{f_+(\xi_n)}{f_-(\xi_n)}\Big| H^{\pm}\right]
\end{aligned}
$$

As a concrete example, if we take

$$
f_{\pm}(\xi) = \frac{1}{\sqrt{2\pi\Delta t\sigma^2}} e^{-(\xi - \Delta t\mu_{\pm})/(2\Delta t\sigma^2)} \tag{5.10}
$$

the above become (in environment $H^{\pm}$)

$$
\begin{aligned}
h(t) &= \pm\frac{(\mu_+ - \mu_-)^2}{2\sigma^2} \\
\rho_{1-c}^2(t) &= (1-c)\frac{(\mu_+ - \mu_-)^2}{\sigma^2} \\
\rho_c^2(t) &= c\frac{(\mu_+ - \mu_-)^2}{\sigma^2}.
\end{aligned} \tag{5.11}
$$

Using $\mu_+ = 1$, $\mu_- = -1$ and $\sigma = \sqrt{2}$, in state $H^{\pm}$ in the continuum limit the belief of agent i evolves according to

$$
dy_i = dt + (\sqrt{2(1-c)}dW_i + \sqrt{2c}dW_c). \tag{5.12}
$$

This provides a correlated-evidence case extension to our previous independent-evidence case SDE, which corresponds to setting $c = 0$ in Eq. (5.12).

We note that dW_j corresponds to private noise, which is generated independently for each agent. The term dW_c is common to all agents. In contrast in the independent-evidence case (or as

$c \to 0^+$), the belief of each agent evolved according to $dy_j = dt + \sqrt{2}dW_j$, and only the drift term was shared between agents.

5.3 Alternative formulations

In the above derivation, we assumed that *all* agents made a common observation with probability c. In an alternative formulation of the model, we could assume that *each* agent makes an observation from the common pool with probability α so that in a group of N agents the probability that all agents make a common observation is α^N. For the case $N = 2$, the resulting SDE is equivalent to Eq. (5.12) with $c = \alpha^2$.

Complications arise in the $N > 2$ case in this formulation. Consider the case $N = 3$. The probability for any two agents to make a common observation is $3\alpha^2(1-\alpha)$, and the probability for all three agents to make a common observation is α^3. Thus, there are effectively *four* sources of common noise: the common observation pool when all agents make a common observation and the pools W_{ij}, which represent the common observation pools when only agents i and j make a common observation, $i, j \in \{1,2,3\}, i \neq j$, while the third agent makes an independent observation. To obtain the correct SDE, it remains to ascertain if the off-diagonal elements of the covariance matrix can be properly chosen to capture the combinatorics of the various sources of common noise.

Another formulation could be to assign each of the N agents a distinct probability c_i, $1 \leq i \leq N$, of making a common observation. This would reflect individual tendencies to rely more on common information ($c_i > 0.5$) or on independent observations ($c_i < 0.5$). It would be interesting to derive the SDE in this case. We hypothesize that first deciders who have a nontrivial bias towards independent or common observations are more likely to be correct than ones for whom $c_i \approx 0.5$.

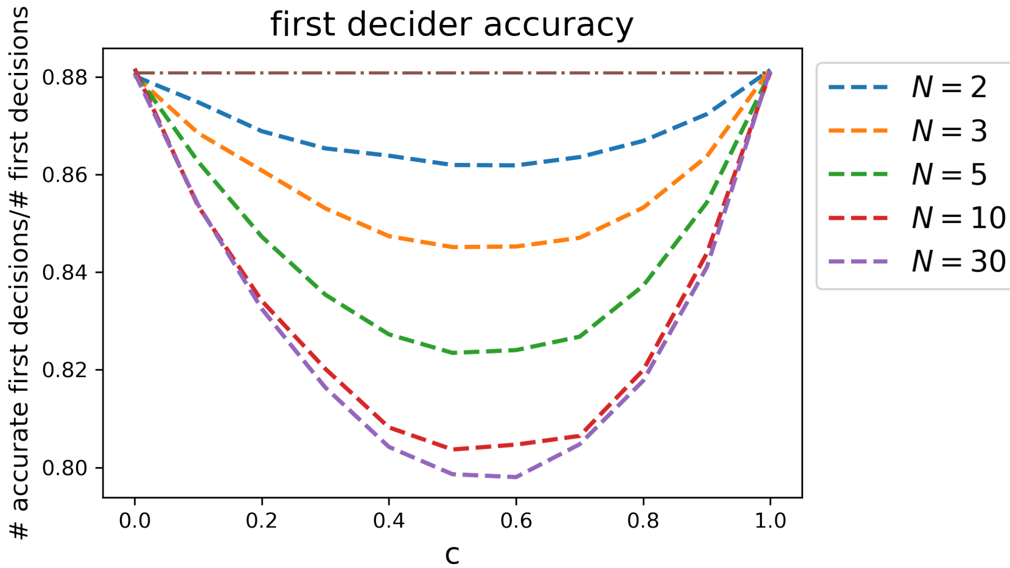

Figure 5.2: Simulation results for first decider accuracy in cliques of size N with threshold $\theta = 2$. The grey dash-dot line gives the expected accuracy for a single decider with threshold θ, $1/(1+e^{-\theta})$. First decider accuracy falls below this benchmark for $0 < c < 1$, when deciders may be expected to receive a mix of independent and common observations.

As we will show subsequently, taking $c \approx 0.5$ roughly minimizes the accuracy of the first agent to decide in our standard model.

5.4 Accuracy discrepancy between the models

In the independent-evidence case, the accuracy of the decision of an agent who makes a choice when their belief reaches a threshold, $\pm\theta$ is given by $P^+(y_i(T_i) = \theta) = 1/(1+e^{-\theta})$ [11]. Here T_i is the first passage time for the belief of agent i, that is, the first time $y_i(t)$ reaches one of the thresholds. Conditioning on the time of an agent's decision does not affect their accuracy, as it is always determined by this simple formula via the log-likelihood ratio. However, in both the discrete and continuous correlated-evidence case, this equality does not always hold since earlier decisions are less likely to be correct. In particular, simulations show that in a network of identical agents receiving correlated evidence, the probability that the first choice is correct depends not only on the threshold, θ, but also the correlation coefficient, c, and the size of the clique N. Figure 5.2 shows simulation results for the accuracy of the first decider in the discrete model when the threshold is set to $\theta = 2$ for various clique sizes N. Considering agents that may draw from a correlated evidence pool reduces the accuracy of the first decider. The first decider's accuracy is convex as a function of $c \in [0,1]$, and attains a minimum close to $c = 0.5$. Moreover, increasing the size of the clique also increases the likelihood that the first decider is wrong. This is in contrast to the independent-evidence case, where the accuracy of the first decider is independent of clique size, and depends only on the size of the threshold, θ.

Hence, for values of c in the interval $(0,1)$ and for cliques of size $N > 1$, the first decider in the correlated-evidence case is less accurate than the first decider in the independent-evidence case. For $c = 0$ and $c = 1$ the accuracy of the first decider in the model with correlated evidence is the same as in the model with independent evidence. This is easy to understand: When $c = 0$ evidence is independent, and we recover the independent evidence model. When $c = 1$ all observers see identical evidence, acting as a single decider.

The observation that accuracy is minimal at an intermediate correlation value was unexpected. Regardless of the value of c, agents in the correlated-evidence case still marginally receive statistically equivalent evidence as in the independent-evidence case. Indeed, if we choose an agent at random, and ask what is the probability that the agent makes a correct choice once their belief reaches threshold, the answer would be the same as in the case when $c = 0$. Each agent, unaware of all other agents, makes decisions that are equivalent to that of an isolated agent, and can tell their accuracy is determined by their threshold. However, someone observing the order in which the decisions are made, but not the evidence itself, can tell which decisions are more accurate. The first and last decider, unaware of this, would still make decisions that, from their perspective, are based on the exact amount of evidence sufficient to reach a decision threshold, and are thus equally likely to be correct.

As a more concrete and potentially experimentally testable example, consider a number of people in separate cubicles gathering evidence, and instructed to make a decision once the total amount of accumulated evidence reaches a threshold. On each timestep the information given is drawn from the same distribution; therefore, as an individual in your own cubicle, what information people in other cubicles are receiving does not impact the observations you see. However, our numerical observations show that the probability that the same observation is received by multiple agents at some timesteps (again, this observation is drawn from the same distribution as the individual observations) affects first decider accuracy. That is, without changing the quality of the observations available, simply making some of the information common makes it more likely that the first person to have enough information to make a decision will make the wrong one.

In the next chapter, we more carefully develop our definitions, show (with some restrictions) the discrete correlated-evidence case model's dependence on the values of c and N, and try to provide an intuitive explanation for this dependence.

Chapter 6

Accuracy analysis for the discrete correlated-evidence model

In the last chapter we observed a decrease in the accuracy of the first decision when evidence is correlated. We now analyze this phenomenon in the context of the discrete version of the correlated-evidence model. In particular, we find the probability that the first decider is correct.

Recall that $y_i(t)$ denotes the belief of agent i at time t, and that we denote by $y_{FD}(t)$ the belief of the first decider at time t. While the identity of the first decider is not predetermined, we could look back at the first decider's trajectory of beliefs $y_{FD}(t)$ for $t < T$ preceding their decisions *after* we know their identity. (If the belief state equals θ for multiple agents at the time of the first decision, we choose one of these agents uniformly, at random and call the chosen agent the first decider.) Recall that $P^{\pm}(\cdot) = P(\cdot|H^{\pm})$ denotes a probability conditioned on the environmental state being $H^{\pm}$. We want to find $P^{\pm}(y_{FD}(T) = \pm\theta)$, where T denotes the time of the first decision. Since the sign of the threshold and the environment match, this quantity corresponds to the probability of a correct first decision. We only need to compute $P^{+}(y_{FD}(T) = \theta)$, since by our assumption of symmetry $P^+(y_{FD}(T)=\theta) = P^-(y_{FD}(T)=-\theta)$, *i.e.* the probability of making a correct choice is the same, regardless of the environmental state.

6.1 Probability of a correct first decision

As in the previous chapter, we assume that the increments in beliefs are equal in magnitude, and thus the belief of each agent evolves on a lattice whose spacing is the belief increment from a single observation. We also assume that the thresholds, $\pm\theta$, are on this lattice. Therefore, either after mapping the belief states to the integers, or assuming $p = q/e$, we can and do assume that the beliefs of individual agents are integer-valued, *i.e.* $y_i(t) \in \mathbb{Z}$ for $1 \leq i \leq N$.

We denote by $\xi(t)$ the vector $\mathbf{y}(t)$ of beliefs of the N agents at time t, extended to include a variable $x(t) \in \{0,1\}$, where $x(t) = 0$ when agents make a common observation at time t and $x(t) = 1$ when the observations at time t are independent. We use the convention that variables that represent a temporal sequence denote the entire sequence when written without an argument. Hence,

$$\xi(t) = (\mathbf{y}(t), x(t)) \text{ and } \xi = \Big(\xi(t)\Big)_{t\in\mathbb{Z}_{\geq 0}},$$

where $\mathbf{y}(t) = (y_1(t), y_2(t), \ldots, y_N(t)) \in \mathbb{Z}^N$ for $t \in \mathbb{Z}_{\geq 0}$ is the vector of belief states of agents in the network at time t. As in earlier chapters, we assume even priors, so that the initial beliefs do not favor either alternative, $\mathbf{y}(0) = (y_1(0), y_2(0), \ldots, y_N(0)) = (0, 0, \ldots, 0)$.

The sample space of the discrete version of the correlated-evidence model consists of sequences, ξ, where each $\xi(t)$ takes values in $\mathbb{Z}^N \times \{0,1\}$. Hereafter, we refer to a single such sequence, ξ, as a *trajectory*. Each trajectory, ξ, thus represents the evolution of the beliefs of all agents, along with a record of whether the observations were independent or common at each timestep. We denote a portion of a trajectory as $\xi(1:k) = \Big(\xi(t)\Big)_{t=1}^{k}$. Similarly, we denote a portion of the belief trajectory by $\mathbf{y}(1:k) = \Big(\mathbf{y}(t)\Big)_{t=1}^{k}$, by $y_i(1:k) = \Big(y_i(t)\Big)_{t=1}^{k}$ the evolution of the belief of agent i from time 1 to time k, and by $x(1:k) = \Big(x(t)\Big)_{t=1}^{k}$ the nature (common or independent) of the first k observations.

Let T be the random variable whose domain is the set of all trajectories and whose value is given by

$$T(\xi) = \max\{\tilde{T} \mid |y_j(t)| < \theta \text{ for all } 1 \leq j \leq N \text{ and } t < \tilde{T}\}$$

if a first decision is made in finite time, and $T(\xi) = \infty$ otherwise[1]. The random variable T is a stopping time that, when finite, equals the integer time of the first decision. Note that $P(T < \theta) = 0$, as it takes at least θ observations to reach a decision.

For a group of N agents with beliefs $\mathbf{y}(t)$ at time t, the probability that agent i decides, and decides correctly at the time of the first decision is

$$P^+(y_i(T) = +\theta),$$

provided we condition on the environmental state H^+. Using the law of total probability, we write this conditional probability of a correct decision by agent i at the time of the first decision as

$$P^+(y_i(T) = +\theta) = \sum_{\tilde{T}=\theta}^{\infty} P^+\Big(y_i(T) = +\theta, T = \tilde{T}\Big). \tag{6.1}$$

We complete this section by expressing (6.1) in terms of random sets.

Define the random set D on the sample space by

$$D(\xi) = \{j \mid y_j\Big(T(\xi)\Big) = \pm\theta\}$$

[1]We have that $P(T < \infty) = 1$ since a biased random walk will almost surely escape the bounded set $\{l \in \mathbb{Z} : -\theta < l < \theta\}$. Any individual agent will therefore almost surely make a decision in finite time.

if $T < \infty$ and $D(\xi) = \emptyset$ if $T = \infty$. That is, D maps each trajectory to the set of agents whose beliefs equal $-\theta$ or θ at the time of the first decision. One can think of D as a set-valued random variable. For each integer $\tilde{T} \geq \theta$, we define a random set $D_{\tilde{T}}$ associated with D by

$$D_{\tilde{T}}(\xi) = \begin{cases} D(\xi), & \text{if } T(\xi) = \tilde{T}, \\ \emptyset, & \text{if } T(\xi) \neq \tilde{T}. \end{cases}$$

Note that agent i is an element of $D_{\tilde{T}}$ if and only if $T = \tilde{T}$ and $y_i(T) = \pm\theta$.

We partition $D(\xi)$ into the subset of agents with belief θ (positive threshold) at decision time $T(\xi)$,

$$D_+(\xi) := \{j \mid y_j(T) = +\theta\} \subseteq D(\xi),$$

and the subset of agents with belief $-\theta$ (negative threshold) at time T,

$$D_-(\xi) := \{j \mid y_j(T) = -\theta\} \subseteq D(\xi).$$

Finally, define $D_{\tilde{T},+}$ and $D_{\tilde{T},-}$ by

$$D_{\tilde{T},\pm}(\xi) = \begin{cases} D_\pm(\xi), & \text{if } T(\xi) = \tilde{T}, \\ \emptyset, & \text{if } T(\xi) \neq \tilde{T}. \end{cases} \tag{6.2}$$

Either $D_+(\xi)$ or $D_-(\xi)$ is empty if $D(\xi)$ contains only wrong or right deciders, respectively.

Expressing the right side of Eq. (6.1) in terms of random sets, we have

$$\sum_{\tilde{T}=\theta}^{\infty} P^+\left(y_i(T) = +\theta, T = \tilde{T}\right) = \sum_{\tilde{T}=\theta}^{\infty} P^+\left(i \in D_{\tilde{T},+}\right).$$

6.2 An expression for $P^{+}(i \in D_{\tilde{T},+})$

We next obtain an expression for the probability $P^{+}(i \in D_{\tilde{T},+})$. We recall from Eq. (6.2) that $P^{+}(i \in D_{\tilde{T},+})$ is the probability that the time of the first decision is $T(\xi) = \tilde{T}$ and the belief of agent i at time T is $y_i(T) = \theta$. Recall that $y_i(1:\tilde{T})$ denotes the portion of the belief trajectory of agent i up to time $\tilde{T}$. We then have

$$P^{+}(i \in D_{\tilde{T},+}) = \sum_{\{y_i(1:\tilde{T})\}} P^{+}(i \in D_{\tilde{T},+} \mid y_i(1:\tilde{T}))P^{+}(y_i(1:\tilde{T})), \tag{6.3}$$

where we sum over all possible finite belief trajectories, $y_i(1:\tilde{T})$.

The probability of a finite belief trajectory for a given agent i is given by the probability of a particular sequence of observations. In particular, let a_i denote the number of observations out of the first $\tilde{T}$ observations made by agent i that *agree* with hypothesis H^{+}. Then

$$P^{+}(y_i(1:\tilde{T})) = p^{a_i} q^{\tilde{T}-a_i}.$$

We further use the finite observation state sequence, $x(1:\tilde{T})$, to write

$$\begin{aligned}
&P^{+}(i \in D_{\tilde{T},+} \mid y_i(1:\tilde{T})) \\
&\qquad = \sum_{\{x(1:\tilde{T})\}} P^{+}(i \in D_{\tilde{T},+} \mid y_i(1:\tilde{T}), x(1:\tilde{T}))P^{+}(x(1:\tilde{T}) \mid y_i(1:\tilde{T})) \\
&\qquad = \sum_{\{x(1:\tilde{T})\}} P^{+}(i \in D_{\tilde{T},+} \mid y_i(1:\tilde{T}), x(1:\tilde{T}))P(x(1:\tilde{T})),
\end{aligned}$$

where we sum over all possible finite observation state trajectories of length $\tilde{T}$. Here $P^{+}(x(1:\tilde{T}) \mid y_i(1:\tilde{T})) = P(x(1:\tilde{T}))$ since the probability of receiving a common observation is independent of environmental state and the belief state of agent i.

We denote the number of independent observations up to time $\tilde{T}$ by $k = k(x) = \sum_{t=1}^{\tilde{T}} x(t)$. Note that k and a_i are functions of the finite trajectory $\xi(1:\tilde{T})$, but we suppress this dependence in both cases. We then have

$$P(x(1:\tilde{T})) = (1-c)^k c^{\tilde{T}-k}. \tag{6.4}$$

6.3 Dependence on the beliefs of a second agent

Thus far, our expressions for the conditional probability that agent i decides and does so correctly at the time of the first decision are not conditioned on the beliefs of other agents. We now condition on the belief trajectories of a second agent. This will eventually allow us to compute the accuracy of the first decision in the context of groups of agents.

Throughout this section we assume $N = 2$ and we use indices i and j for the two agents. Let R_j be the sequence of *independent* observations made by agent $j \neq i$. (The sequence R_j can be determined given knowledge of y_j and x.) For any countable set A_j of sequences of independent observations, we have

$$P^+(i \in D_{\tilde{T},+}, R_j \in A_j \mid y_i, x) = \sum_{R_j \in A_j} P^+(i \in D_{\tilde{T},+} \mid R_j, y_i, x) P^+(R_j \mid y_i, x). \tag{6.5}$$

When performing computations that involve a fixed decision time $\tilde{T}$, we work with finite trajectories (from time 1 to time $\tilde{T}$), though we leave this implicit in the notation. For instance, only the finite sequence $R_j(1:\tilde{T})$ matters in Eq. (6.5). If we let a_{R_j} be the number of independent observations of agent j up to time $\tilde{T}$ that agree with option H^+, then

$$P^+(R_j(1:\tilde{T}) \mid y_i, x) = P^+(R_j(1:\tilde{T}) \mid x) = p^{a_{R_j}} q^{k(x)-a_{R_j}}.$$

Here $k(x)$ is again the number of independent observations.

The conditional probability $P^+(i \in D_{\tilde{T},+} \mid R_j, y_i, x)$ on the right side of Eq. (6.5) takes value zero or one. Thus, computing the joint probability is essentially a counting problem:

$$\begin{aligned}
&P^+(i \in D_{\tilde{T},+}, R_j \in A_j) = \sum_{\{y_i\}} P^+(i \in D_{\tilde{T},+}, R_j \in A_j \mid y_i) P^+(y_i) \\
&= \sum_{\{y_i\}} P^+(y_i) \left[\sum_{\{x\}} P^+(i \in D_{\tilde{T},+}, R_j \in A_j \mid y_i, x) P(x) \right] \\
&= \sum_{\{y_i\}} P^+(y_i) \left[\sum_{\{x\}} P(x) \left(\sum_{R_j \in A_j} P^+(i \in D_{\tilde{T},+} \mid R_j, y_i, x) P^+(R_j \mid y_i, x) \right) \right] \\
&= \sum_{\{y_i\}} p^{a_i} q^{\tilde{T}-a_i} \left[\sum_{\{x\}} (1-c)^{k(x)} c^{\tilde{T}-k(x)} \left(\sum_{R_j \in A_j} P^+(i \in D_{\tilde{T},+} \mid R_j, y_i, x) p^{a_{R_j}} q^{k(x)-a_{R_j}} \right) \right],
\end{aligned} \tag{6.6}$$

where, again, a_i is the number of observations (belief updates) consistent with option H^+ in the belief trajectory y_i, and a_{R_j} is the number of observations consistent with option H^+ in the trajectory of independent observations, R_j, of agent j. An expression similar to Eq. (6.6) can be obtained for the joint probability of a decision by agent i, and belief sequences of more than two agents.

6.4 Evaluating $P^+(i \in D_{\tilde{T},+})$ when $\tilde{T} = \theta$

The counting problem encoded by Eq. (6.6) can be complicated when $\tilde{T} > \theta$. When $\tilde{T} = \theta$, however, simplifications become possible. In this section, we continue to assume that $N = 2$ and we set $\tilde{T} = \theta$.

Eq. (6.6) simplifies in several ways. First, when $\tilde{T} = \theta$ only one finite belief trajectory $y_i(1:\theta)$ contributes to the sum, namely the trajectory with all observations in agreement with H^+, so that $a_i = \theta$. Second, since $T = \theta$, we must sum over all possible $x(1:\theta)$. We can do so by summing over the number of independent observations made between times 1 and θ. These simplifications yield

$$P^+(i \in D_{\theta,+}, R_j \in A_j)$$
$$= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{R_j \in A_j} P^+(i \in D_{\theta,+} \mid R_j, \hat{y}_i, k) p^{a_{R_j}} q^{k-a_{R_j}} \right) \right], \tag{6.7}$$

where $\hat{y}_i$ refers to the finite belief trajectory from time 1 to time θ with $a_i = \theta$.

We can rewrite the second sum as a sum over the values of a variable $R_{+,j}(\theta, y_j(\theta)) = \theta - y_j(\theta)$, which gives the distance of the belief of agent j from the belief of agent i at decision time. (We assume $y_i(\theta) = \theta$.) Note that $R_{+,j}$ can take even integer values r_+ between 0 and 2θ. Let $W_j \subset \{0, 2, \ldots, 2\theta\}$. The joint probability in Eq. (6.7) now has the equivalent form

$$P^+(i \in D_{\theta,+}, R_{+,j} \in W_j)$$
$$= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{\substack{r_+ \in W_j \\ k-r_+/2 \geq 0}} \frac{k! p^{k-r_+/2} q^{r_+/2}}{(k-r_+/2)!(r_+/2)!} \right) \right]. \tag{6.8}$$

Eq. (6.7) and Eq. (6.8) are equivalent when the set A_j is identified with the set of trajectories for which $R_+ \in W_j$.

Using the expression in Eq. (6.8), it is possible to express restrictions on the belief of the second agent by choosing W_j. For example, $P^+(i \in D_{\theta,+}, j \notin D_{\theta,+})$ adds only the requirement that $y_j(\theta) < \theta$, which is equilvalent to $R_{+,j} \in \{2, 4, \ldots, 2\theta\}$, thus giving

$$P^+(i \in D_{\theta,+}, j \notin D_{\theta,+})$$
$$= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{\substack{r_+ \in \{2,4,\ldots,2\theta\} \\ k-r_+/2 \geq 0}} \frac{k! p^{k-r_+/2} q^{r_+/2}}{(k-r_+/2)!(r_+/2)!} \right) \right]. \tag{6.9}$$

Similarly, we also have

$$\begin{aligned}
&P^+(i \in D_{\theta,+}, j \in D_{\theta,+}) \\
&= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{r_+=0} \frac{k! p^{k-r_+/2} q^{r_+/2}}{(k-r_+/2)!(r_+/2)!} \right) \right] \\
&= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} p^k \right) \Bigg],
\end{aligned} \tag{6.10}$$

$$\begin{aligned}
&P^+(i \in D_{\theta,+}, j \notin D_\theta) \\
&= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{\substack{r_+ \in \{2,4,\ldots,2\theta-2\} \\ k-r_+/2 \geq 0}} \frac{k! p^{k-r_+/2} q^{r_+/2}}{(k-r_+/2)!(r_+/2)!} \right) \right],
\end{aligned} \tag{6.11}$$

and

$$\begin{aligned}
&P^+(i \in D_{\theta,+}, j \in D_{\theta,-}) \\
&= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{r_+=2\theta, k=\theta} \frac{k! p^{k-r_+/2} q^{r_+/2}}{(k-r_+/2)!(r_+/2)!} \right) \right] \\
&= p^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} q^\theta \right) \Bigg].
\end{aligned} \tag{6.12}$$

If we are instead interested in $P^+(i \in D_{\theta,-}, R_j \in A_j)$, we may obtain it analogously by reversing our p and q values and using $R_{-,j}(\theta, y_j(\theta)) = y_j(\theta) - (-\theta)$ instead of $R_{+,j}$:

$$P^+(i \in D_{\theta,-}) = q^\theta \left[\sum_{k=0}^{\theta} \frac{\theta!}{k!(\theta-k)!}(1-c)^k c^{\theta-k} \left(\sum_{\substack{r_- \in \{0,2,\ldots,2\theta\} \\ k-r_-/2 \geq 0}} \frac{k! q^{k-r_-/2} p^{r_-/2}}{(k-r_-/2)!(r_-/2)!} \right) \right]. \tag{6.13}$$

6.5 $P^+(i \in D_{\tilde{T},+})$ when $\tilde{T} = \theta$ for arbitrary N

For $\tilde{T} = \theta$ and $N = 2$, we have obtained expressions for the probability that agent i decides at time θ and does so correctly by conditioning on the beliefs of the second agent. When $N > 2$, we can obtain analogous expressions for agent i by conditioning on the beliefs of the other $N-1$ agents.

We previously introduced the distance $R_{+,j}$ of a single agent j from the deciding agent i and noted that this random variable takes a finite set of values $r_+ \in \{0,2,\ldots,2k\}$ with $2k \leq 2\theta$. For $N > 2$ agents, we define the vector $\mathbf{R}_{+,\hat{\imath}} = (R_{+,1},\ldots,R_{+,i-1},R_{+,i+1},\ldots,R_{+,N})$. Each $R_{+,j}$ can then take values $r_{+,j} \in \{0,2,\ldots,2k\}$ with $2k \leq 2\theta$. As in the previous section, when there are k independent observations, only the values $k - r_{+,j}/2 \geq 0$ are allowable. If we let $W_{\hat{\imath}}$ be some set of possible values of $\mathbf{R}_{+,\hat{\imath}}$, Eq. (6.8) generalizes to

$$P^+(i \in D_{\theta,+}, \mathbf{R}_{+,\hat{\imath}} \in W_{\hat{\imath}}) = p^\theta \Big[\sum_{k=0}^{\theta} \frac{\theta!(1-c)^k c^{\theta-k}}{k!(\theta-k)!} \Big(\sum_{\substack{\mathbf{r}_{+,\hat{\imath}} \in W_{\hat{\imath}} \\ k - r_{+,j}/2 \geq 0 \,\forall j \neq i}} \Big[\prod_{\mathbf{r}_{+,\hat{\imath}}} \frac{k! p^{k-r_{+,j}/2} q^{r_{+,j}/2}}{(k-r_{+,j}/2)!(r_{+,j}/2)!} \Big] \Big) \Big]. \tag{6.14}$$

The product is over the entries of the vector $\mathbf{r}_{+,\hat{\imath}}$.

6.6 From $P^+(i \in D_{\tilde{T},+})$ to $P^+(y_{FD} = +\theta)$ for $N = 2$

Let the first decider (FD) be a random variable whose argument is the decision set D of a trajectory ξ and whose value is a single agent selected with equal probability from that decision set. Then the probability that the first decider is correct (has reached the threshold consistent with the environmental state $H^{\pm}$) is given by

$$\begin{aligned}
&P^{+}(y_{FD}=+\theta)\\
&\quad=\sum_{\{\xi\}}P^{+}(y_{FD}=+\theta,D(\xi))\\
&\quad=\sum_{\{\xi\}}\left(\sum_{i}P^{+}\Big((y_{FD}=+\theta,D(\xi)),FD(D(\xi))=i\Big)\right)\\
&\quad=\sum_{\{\xi\}}\left(\sum_{i}P^{+}\Big(y_{FD}=+\theta,D(\xi)\mid FD(D(\xi))=i\Big)P^{+}\Big(FD(D(\xi))=i\Big)\right) \qquad (6.15)\\
&\quad=\sum_{\{\xi\}}\left(\sum_{i}P^{+}\Big(y_{i}=+\theta,D(\xi)\Big)P^{+}\Big(FD(D(\xi))=i\Big)\right)\\
&\quad=\sum_{\{\xi\}}\left(\sum_{i}P^{+}\Big(i\in D_{+}(\xi),D(\xi)\Big)P^{+}\Big(FD(D(\xi))=i\Big)\right).
\end{aligned}$$

Exchanging the order of summation in Eq. (6.15), we can use

$$\begin{aligned}
P^{+}(y_{FD}=+\theta)&=\sum_{i}\sum_{\{\xi\}}P^{+}\Big(i\in D_{+}(\xi),D(\xi)\Big)P^{+}\Big(FD(D(\xi))=i\Big)\\
&=\sum_{i}\sum_{D,D_{+}\subset D}P^{+}\Big(i\in D_{+},D\Big)P^{+}\Big(FD(D)=i\Big) \qquad (6.16)\\
&=\sum_{i}\sum_{\substack{D:i\in D,\\ D_{+}\subset D}}\frac{1}{|D|}P^{+}\Big(i\in D_{+},D\Big).
\end{aligned}$$

The interior sum is over all possible decision sets (subsets of $\{1,2,\ldots,N\}$) that contain i and all possible subsets of each of these decision sets.

In the two-agent case $(N=2)$ our potential values for D are $\{1\}$, $\{2\}$, and $\{1,2\}$ so that we have (writing only terms with non-zero probabilities):

$$
\begin{aligned}
P^{+}(y_{FD}=\theta) &= P^{+}(1\in D_{+},2\notin D)\times 1 + P^{+}(2\in D_{+},1\notin D)\times 1 \\
&\quad + P^{+}(1\in D_{+},2\in D)\left(\frac{1}{2}\right) + P^{+}(2\in D_{+},1\in D)\left(\frac{1}{2}\right) \\
&= \left[P^{+}(1\in D_{+},2\notin D) + \frac{1}{2}P^{+}(1\in D_{+},2\in D)\right] \\
&\quad + \left[P^{+}(2\in D_{+},1\notin D) + \frac{1}{2}P^{+}(2\in D_{+},1\in D)\right].
\end{aligned}
\tag{6.17}
$$

where the second equality is a rearrangement of the first. The terms in the two square brackets in Eq. (6.17) are equal by symmetry, and hence

$$
\begin{aligned}
P^{+}(y_{FD}=+\theta) &= 2P^{+}(1\in D_{+},2\notin D) + 2\left(\frac{1}{2}\right)P^{+}(1\in D_{+},2\in D) \\
&= 2P^{+}(1\in D_{+},2\notin D) + P^{+}(1\in D_{+},2\in D_{-}) \\
&\qquad + P^{+}(1\in D_{+},2\in D_{+}),
\end{aligned}
$$

where the second line breaks the last term of the first line into two parts using $D = D_{+}\cup D_{-}$.

Combining this with the partition of the sample space over values of the first decision time, T,

$$
\begin{aligned}
P^{+}(y_{FD}=+\theta) = \sum_{\tilde{T}} \Big(& P^{+}(1\in D_{\tilde{T},+},2\in D_{\tilde{T},+}) + 2P^{+}(1\in D_{\tilde{T},+},2\notin D_{\tilde{T}}) \\
& + P^{+}(1\in D_{\tilde{T},+},2\in D_{\tilde{T},-})\Big).
\end{aligned}
\tag{6.18}
$$

When we fix the decision time at $T=\theta$ as we did in the previous section, we are computing the conditional probability $P^{+}(y_{FD}=+\theta \mid T=\theta)$. The probability that a first decision occurs at

$T = \theta$ is given by

$$
\begin{aligned}
&P^+(T=\theta) \\
&\quad = \Big(P^+(1\in D_{\theta,+}, 2\in D_{\theta,+}) + 2P^+(1\in D_{\theta,+}, 2\notin D_\theta) + P^+(1\in D_{\theta,+}, 2\in D_{\theta,-})\Big) \\
&\qquad + \Big(P^+(1\in D_{\theta,-}, 2\in D_{\theta,-}) + 2P^+(1\in D_{\theta,-}, 2\notin D_\theta) + P^+(1\in D_{\theta,-}, 2\in D_{\theta,+})\Big)
\end{aligned}
$$

where the first line on the right-hand side gives the probability of a correct first decision at time $T = \theta$ and the second line on the right-hand side gives the probability of a wrong first decision at time $T = \theta$. Then we have

$$
\begin{aligned}
P^+(y_{FD} = +\theta \mid T=\theta) &= \frac{P^+(y_{FD}=\theta, T=\theta)}{P^+(T=\theta)} \\
&= \frac{P^+(1,2\in D_{\theta,+}) + 2P^+(1\in D_{\theta,+}, 2\notin D_\theta) + P^+(1\in D_{\theta,+}, 2\in D_{\theta,-})}{P^+(T=\theta)} \\
&= \frac{1}{1+\frac{P^+(1,2\in D_{\theta,+})+2P^+(1\in D_{\theta,+},2\notin D_\theta)+P^+(1\in D_{\theta,+},2\in D_{\theta,-})}{P^+(1,2\in D_{\theta,-})+2P^+(1\in D_{\theta,-},2\notin D_\theta)+P^+(1\in D_{\theta,-},2\in D_{\theta,+})}} \qquad (6.19)
\end{aligned}
$$

where the third line is obtained by factoring out the numerator in the second.

Figure 6.1 shows the conditional probability of a correct decision given by Eq. (6.19) compared to that obtained using numerical simulations. We see that even when conditioning on decision time, the dip in accuracy for middling degrees of correlation is evident.

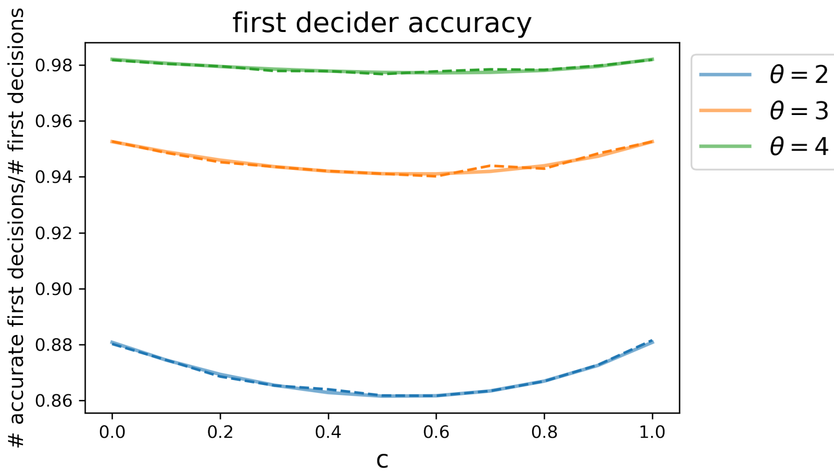

Figure 6.1: Accuracy of first decider for the correlated-evidence case using Eq. (6.19) for $\theta = 2$, 3, 4. The solid line is the probability of an accurate first decision given by Eq. (6.19) (which is exact) and the dashed lines are simulation results for the correlated-evidence case conditioned on a first decision at the earliest possible time $T = \theta$.

6.7 From $P^+(i \in D_{\tilde{T},+})$ to $P^+(y_{FD} = +\theta)$ for arbitrary N

To extend our formula for $P^+(y_{FD} = \theta)$ to allow for arbitrary group size N, we begin with Eq. (6.16):

$$\begin{aligned} P^+(y_{FD} = \theta) &= \sum_i \sum_{D, D_+ \subset D} P^+\Big(i \in D_+, D\Big) P^+\Big(FD(D) = i\Big) \\ &= \sum_i \sum_{\substack{D: i \in D, \\ D_+ \subset D}} \frac{1}{|D|} P^+\Big(i \in D_+, D\Big). \end{aligned}$$

By symmetry, the values of the outer summation over i will all be equal so that instead have

$$P^+(y_{FD} = \theta) = N \sum_{\substack{D: i \in D, \\ D_+ \subset D}} \frac{1}{|D|} P^+\Big(i \in D_+, D\Big) \tag{6.20}$$

For the inner probability $P^+\Big(i \in D_+, D\Big)$, we use Eq. 6.14 where we choose $W_{\hat{i}}$ to be the set of all vectors $\mathbf{R}_{+,\hat{i}}$ such that the decision set is D and $i \in D_+$. Figure 6.2 gives the accuracy of this

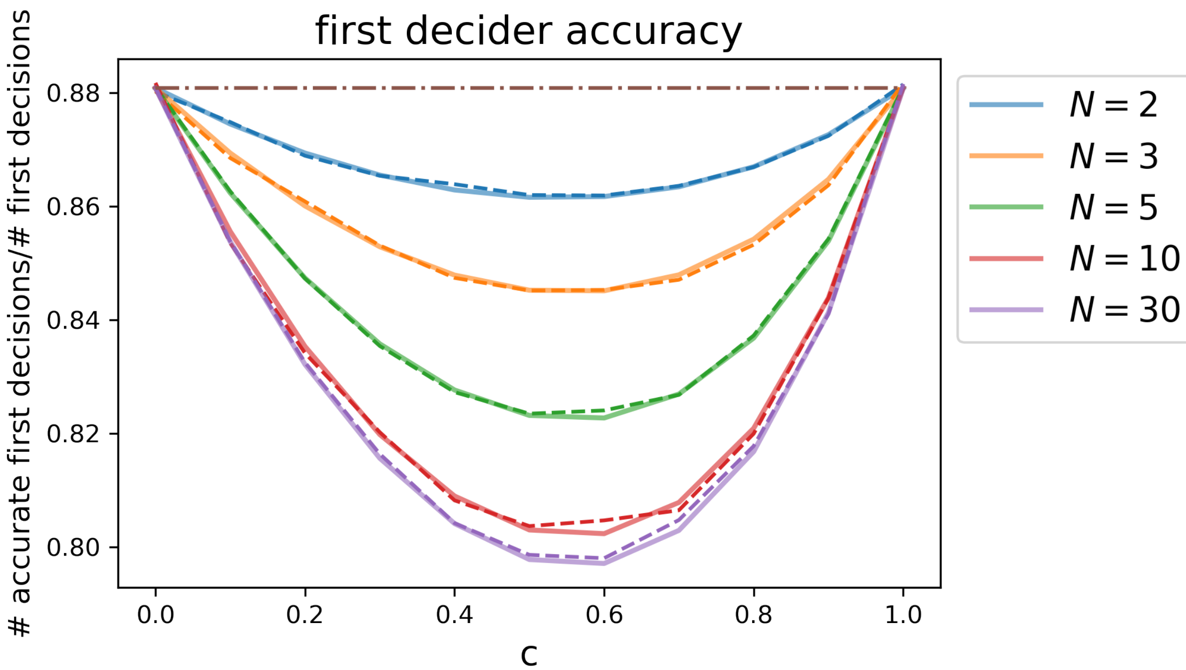

Figure 6.2: Accuracy of first decider of a group of N agents for threshold $\theta = 2$ for first decisions occurring at the earliest possible decision time $T = \theta$. Dashed lines show simulation results and solid lines show theory according to Eq. (6.20).

formula for threshold $\theta = 2$ and various values of N.

6.8 Approximating $P^+(y_{FD} = \theta)$ for large N

The formulas provided in the previous sections, while accurate, are also prohibitively complicated as T grows large. This complication stems from the difficulty of enumerating the set of desired values of x to provide an accurate formula for $P(k(x)|T)$ for values of T larger than θ. (See Eq. (6.7) for the solution to this enumeration when $T = \theta$.) Accordingly, if we wish to extend an accuracy formula to acquire an expected first decider accuracy in the continuous case we require a more tractable approximation. We hope that the following will provide both this tractable approximation and some intuition as to the mechanics behind the dip in accuracy for intermediate values of c.

To justify the simplifying assumptions we will make in this section, we observe that when first decisions occur at a time later than $T = \theta$, some portion of the first decider's observations were in conflict: some contributed to the first decision, and others amounted to dithering.

When N is large, often the majority (perhaps even all) of the first decider's independent observations are in the direction of the first decision; thus, we assume that all of the first decider's independent observations *contribute* to the decision. In some cases, the time of the first decision is much later than the earliest possible time ($T = \theta$). This may often occur if most of the observations are correlated, so the large group is less likely to contain a single individual with a rare stream of observations with the same polarity. Thus, given a first decider's decision time T, we conjecture that the $T - \theta$ beyond the minimal number $T = \theta$ arise due to correlated observations.

We find that when the number of contributing common observations is positive, the contributing common observations completely determine the direction (and therefore the accuracy) of the first decision.

As illustration, consider a trial of the case $\theta = 2$, and suppose that the first observation is common and in the positive direction. Thus in this trial, at time $t = 1$, all agents have belief equal to 1.

The first important point is that, no matter what observations agents receive next, no agent can make an incorrect first decision after the next (second) observation. Even if every agent were to receive a negative observation in the next timestep (either through a common observation or N independent negative observations), their belief would return to 0 rather than reaching the negative threshold.

The second point is that if even a single agent receives a positive observation in the second timestep, their belief will reach the positive threshold and a correct first decision will be made at time $T = 2$. However, if the second observation is common (correlated across agents), with probability q, all agents' beliefs return to 0. On the other hand, if the second observation is independent, the probability that at least 1 out of N agents will receive a positive observation is $(1 - q^N)$. Thus, as N becomes large, it is highly likely the first decision will occur at the second timestep (probability$\rightarrow$ 1 as $N \rightarrow \infty$).

By the same logic as above, if we had supposed that the first observation was common and negative and the second observation was independent, the probability of at least 1 agent receiving a negative independent observation on the second timestep would be $(1 - p^N)$, which also approaches 1 as N becomes large.

The third point is that in this example we had one contributing common observation (the first). If our second observation is independent, we also have one contributing independent observation. However, of the two, only the common observation determines the direction of the decision: for large enough N, for both positive and negative common observations the probability of an according independent observation approaches 1. Thus, the direction (and accuracy) of the first decision is completely determined by the direction of a single observation and has the probability of being accurate we would expect from a decision made from a single observation. Under these conditions (1 contributing common decision), we say that the first decision had an effective *pseudothreshold* of $\tilde{\theta} = 1$.

This same logic can be applied when the first observation is independent and the second observation is common. In that case, the common observation still determines the direction of the first decisions, and in the large N limit it is very likely at least one agent will receive an according independent observation first.

We describe a pseudothreshold as a threshold which approximates the amount of information which determines the accuracy of the first decision under a given first decision time and given number of independent and common observations. We find that for fixed threshold θ and decision time T, we can give adequate approximating pseudothresholds as a function of the number of contributing common observations.

When the number of contributing common observations is positive, we have a pseudothreshold equal to the number of contributing common observations. When the number of contributing common observations is 0, the direction of the decision is instead dominated by the number of contributing independent observations.

Heretofore we have assumed that the probability p of drawing a positive observation and the probability q of drawing a negative observation are related with $q = p/e$ so that our updates are of size $\log p/q = 1$. We can expand the idea of pseudothresholds to encompass versions of the model in which the update is of size υ by instead using the relationship

$$q = \frac{p}{e^{\upsilon}} \; ; \; p + q = 1$$

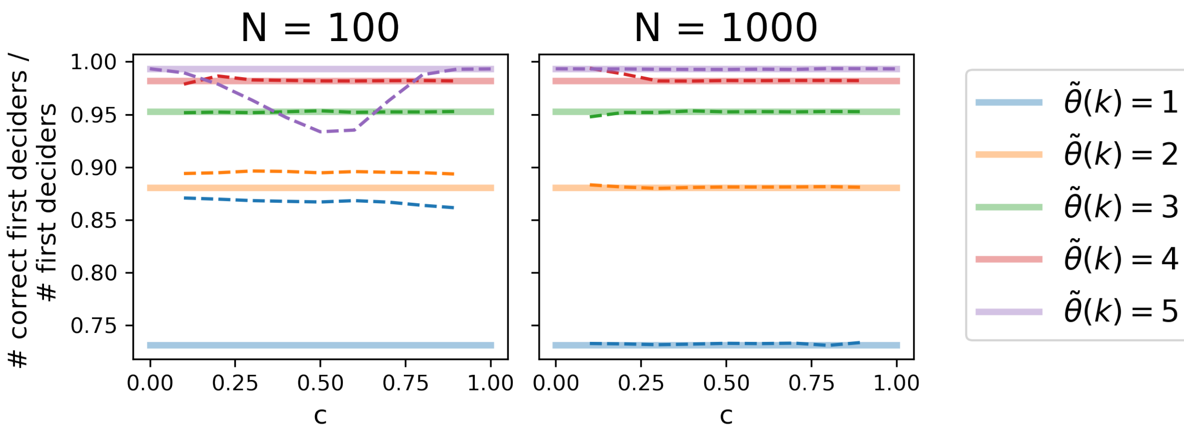

Figure 6.3: Accuracy conditioned on number of independent observations k calculated using a corresponding pseudothreshold $\theta(k)$ (solid lines) vs simulation accuracy (dashed lines) for various k values. The accuracy given by $\tilde{\theta}(k)$ performs poorly for smaller N values (left), but performs well when N is large enough that the assumptions behind the approximation hold (right). Actual threshold $\theta = 5$, update size $\upsilon = 1$. Simulation accuracy is computed by dividing the number of accurate trials with a given k value by the total number of trials with the same given k value.

so that

$$p = \frac{1}{1 + e^{-\upsilon}}$$

and

$$\log \frac{p}{q} = \upsilon. \tag{6.21}$$

To find the number of contributing common observations, we subtract the difference between the actual first decision time T and the minimum first decision time $T_{min} = \frac{\theta}{\upsilon}$ from the actual number of common observations $T - k$ (We recall that the variable k gives the number of independent observations in any particular trial.) Then the size of the pseudothreshold is determined by multiplying either the number of contributing common observations $\left((T-k) - (T - \frac{\theta}{\upsilon})\right)$ or the number of contributing independent observations (k) by the update size υ.

Accordingly, we have

$$\tilde{\theta}(k) = \begin{cases} \theta - k\upsilon & \frac{\theta}{\upsilon} > k \\ k\upsilon & \frac{\theta}{\upsilon} = k \\ 0 & else \end{cases}$$

so that the value of the pseudothreshold $\tilde{\theta}$ may be given as a function of the actual threshold θ, the update size υ, and the number of independent observations k. While the value of $\tilde{\theta}$ is only an approximation, we find that it holds well when N is sufficiently large. (See Figure 6.3 for comparison of simulation results and those given using the pseudothreshold.) Furthermore, the fraction of trials for which $\tilde{\theta} = 0$ is acceptably small.

Since for any particular case θ and υ are constant, we may consider the value of $\tilde{\theta}$ to be essentially a function of k so that for arbitrary update size υ, we have

$$\begin{aligned} P^{+}(y_{FD} = \theta) &= \sum_{k} P^{+}(y_{FD} = \theta|k)P(k) \\ &= \sum_{k} P^{+}(y_{FD} = \theta|k)\left(\sum_{T} P(k|T)P(T)\right. \\ &= \sum_{T} P(T)\sum_{k} P^{+}(y_{FD} = \theta|k)P(k|T)\Big) \\ &\approx \sum_{T} P(T)\left(\sum_{k=0}^{\theta/\upsilon} P^{+}(y_{FD} = \theta|k)P(k|T)\right) \\ &\approx \sum_{T} P(T)\left(\sum_{k=0}^{\theta/\upsilon} P^{+}(y_{FD} = \theta|\tilde{\theta}(k))P(k|T)\right) \\ &= \sum_{T} P(T)\left(\sum_{k=0}^{\theta/\upsilon} \frac{1}{1+e^{-\tilde{\theta}(k)}} P(k|T)\right). \end{aligned} \tag{6.22}$$

We use numerical approximations for $P(T)$. To give $P(k|T)$, we begin with Eq. (6.4) which gives the probability of receiving the desired number of independent and common observations in a specified order. However, as in previous sections, giving the exact number of possible orderings when $T > T_{min}$ is combinatorially complicated. Instead, we seek a reasonable approximation.

In constructing our pseudothresholds we have assumed that the majority of non-contributing common observations precede the majority of independent observations. Accordingly, we approximate by assuming some fraction $(1-c)$ of the non-contributing common observations are made first. This fraction corresponds to the expected fraction of independent observations. Using these assumptions we approximate $P(k|T)$ with

$$\begin{aligned} P(k|T) &\approx P\left(k, x(i) = 0 | T, i = 1..\left(T - \frac{\theta}{\upsilon}\right)(1-c)\right) \\ &= (1-c)^k c^{\bar{t}-k} \frac{\bar{t}!}{k!(\bar{t}-k)!} \end{aligned} \tag{6.23}$$

where $\bar{t}$ is defined as

$$\begin{aligned} \bar{t} &= T - \left(T - \frac{\theta}{\upsilon}\right)(1-c) \\ &= \frac{\theta}{\upsilon} - \left(T - \frac{\theta}{\upsilon}\right)c \\ &= (1-c)\frac{\theta}{\upsilon} - cT. \end{aligned}$$

The fraction $\frac{\theta}{\upsilon}$ is the actual threshold divided by the update size and gives the minimum number of timesteps required to reach the actual threshold which is the minimum possible time of the first decision.

This method of approximating is inexact and results in a discrepancy between the approximate and simulation accuracy particularly for cases with large values of $\frac{\theta}{\upsilon}$. Figure 6.4 compares the results of Eq. (6.22) when Eq. (6.23) is used for $P(k|T)$ and when the value of $P(k|T)$ is taken from numerical results. While inexact, the approximation has the virtue of being sufficiently tractable that future work may attempt to use it as a bridge to an accuracy approximation for the continuous case.

The good performance of Eq. (6.22) when $P(k|T)$ is not approximated roughly suggests that

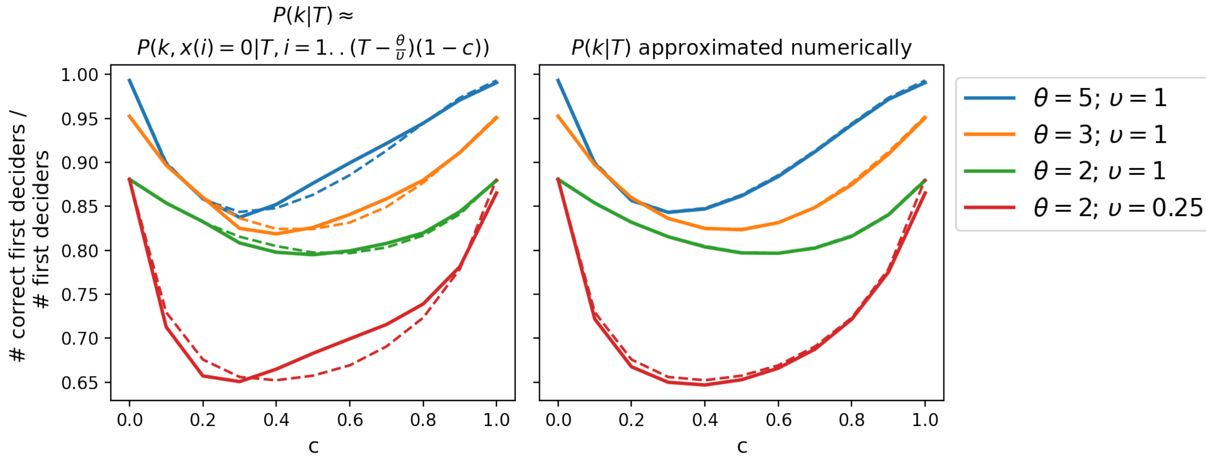

Figure 6.4: Comparison of accuracy given by Eq. 6.22 and simulation results for various thresholds θ and update sizes υ. Approximating $P(k|T)$ with Eq. (6.23) (left) is less exact when the minimum number of timesteps is large. However, using an exact value obtained numerically (right) shows that the pseudothreshold approximation in Eq. (6.22) can perform well. Simulations use clique size $N = 1000$.

the pseudothreshold explanation for the dip in accuracy is correct: The lower accuracy results from a diminution in the amount of evidence that controls the direction of the first decision when a mix of common and independent observations contribute to the first decision.

Chapter 7

Conclusions and future directions

In the earlier chapters we explored the ways temporally structured private and social information shape collective decisions. To address this topic, we considered a network of rational agents who independently accumulate private evidence that triggers a decision upon reaching a threshold. We saw that when seen by the whole network, the first agent's choice initiated a wave of new decisions but later decisions had less impact. The overall probability of a randomly selected agent in such groups making a correct decision was bounded from above because of the impact of the first decider's choice.

In heterogeneous networks, the first decisions were made quickly by impulsive individuals who needed little evidence to make a choice. However, these early decision, even when wrong, revealed the correct options to nearly everyone else. We conclude that groups comprised of diverse individuals can make more efficient decisions than homogeneous ones.

However, when making decisions, we often rely on a mix of information that we have acquired individually and information that is commonly available. Many of the previous studies in this area focus on how well individuals incorporate and account for the effects of correlated evidence.

We chose to neglect the effect of social information exchange and focused instead on a more fundamental question: How does the simple fact of information being individual or common affect the quality of decisions?

To find an answer, we assumed that multiple non-interacting agents make observations and decide between two options when they gathered sufficient information to reach one of two symmetric thresholds. Some observations are made in common by all agents and some privately by each agent. We found that the presence of a mix of common and individual observations results in a decrease in the probability that the first agents reaching threshold makes the correct choice compared to when all observations are private or when all observations are common.

This phenomenon appeared even when private and common observations were equally informative. Therefore, it is only the order of a decision that impacts its accuracy. We explained this counterintuitive observation, and conclude that access to common information decreases accuracy for those whose early private information coincides with the common information.

Future work will include refining the approximation given in Eqs. (6.22), (6.23) for use in limiting to a continuous accuracy equation. It may be that the group size N required to satisfy the assumptions of the approximation grows as the update size υ decreases. If this is the case, it may be possible to acquire values of $\tilde{\theta}$ for trials in which not all independent observations contribute to the first decision, allowing us to relax some of the assumptions while retaining the main idea behind the approximation. It may also be possible to provide a more exact but still tractable approximation for $P(k|T)$, which would enhance the accuracy of the approximation.

In either case, it is desirable to acquire some lower bound on the value of N required for the accuracy of the approximation in Eq. (6.22). It may be that this lower bound will depend on update size υ, perhaps as a proxy for the minimum decision time $T_{min} = \frac{\theta}{\upsilon}$.

Some additional phenomena surfaced over the course of these projects that invite further investigation. In this dissertation, we focused on the unexpected decrease in first decider accuracy as a result of a mix of common and independent observation. However, this is merely a piece of a larger set of accuracy anomalies.

If group members are permitted to collect information, common and individual, until they each reach threshold, numerical simulations show that, in expectation, accuracy increases with decider order when the minimum number of timesteps is greater than 2. That is, the second decider is expected to be more accurate that the first; the third more accurate than the second, and so on. (See Figure 7.1)

While earlier deciders are less accurate than a single agent with the same threshold, the accuracy of later deciders exceeds this benchmark. We would like to extend the intuition behind the lowered first decider accuracy given with the pseudothreshold approximation (Eq. (6.22)) to account for the increasing expected accuracy of later deciders.

As an additional curiosity, when the minimum number of timesteps is $T_{min} = \theta/\upsilon = 2$, it is a middle decider that has the greatest accuracy. (Compare Figure 7.2 with Figure 7.1.)

Another aspect of this phenomenon is its dependence on clique size N. According to simulation results, as clique size increases the accuracy of the first decider quickly reaches a lower bound. However, the accuracy of the last decider continues to increase. As one might intuitively expect, the time of the first decision quickly approaches the minimum necessary time while the time required for the last decider to reach threshold grows exponentially. (See Figure 7.3.)

Numerical results show a somewhat rosier picture for intermediate deciders. If we follow the behavior of the middle decider (decider $N/2$), we find that the average time of the $(\frac{N}{2})^{th}$ agent's decision seems to plateau at a value quite closer to the first decider's average time which is at the earliest possible time. The $(\frac{N}{2})^{th}$ agent's accuracy, however, rapidly aligns with the accuracy of

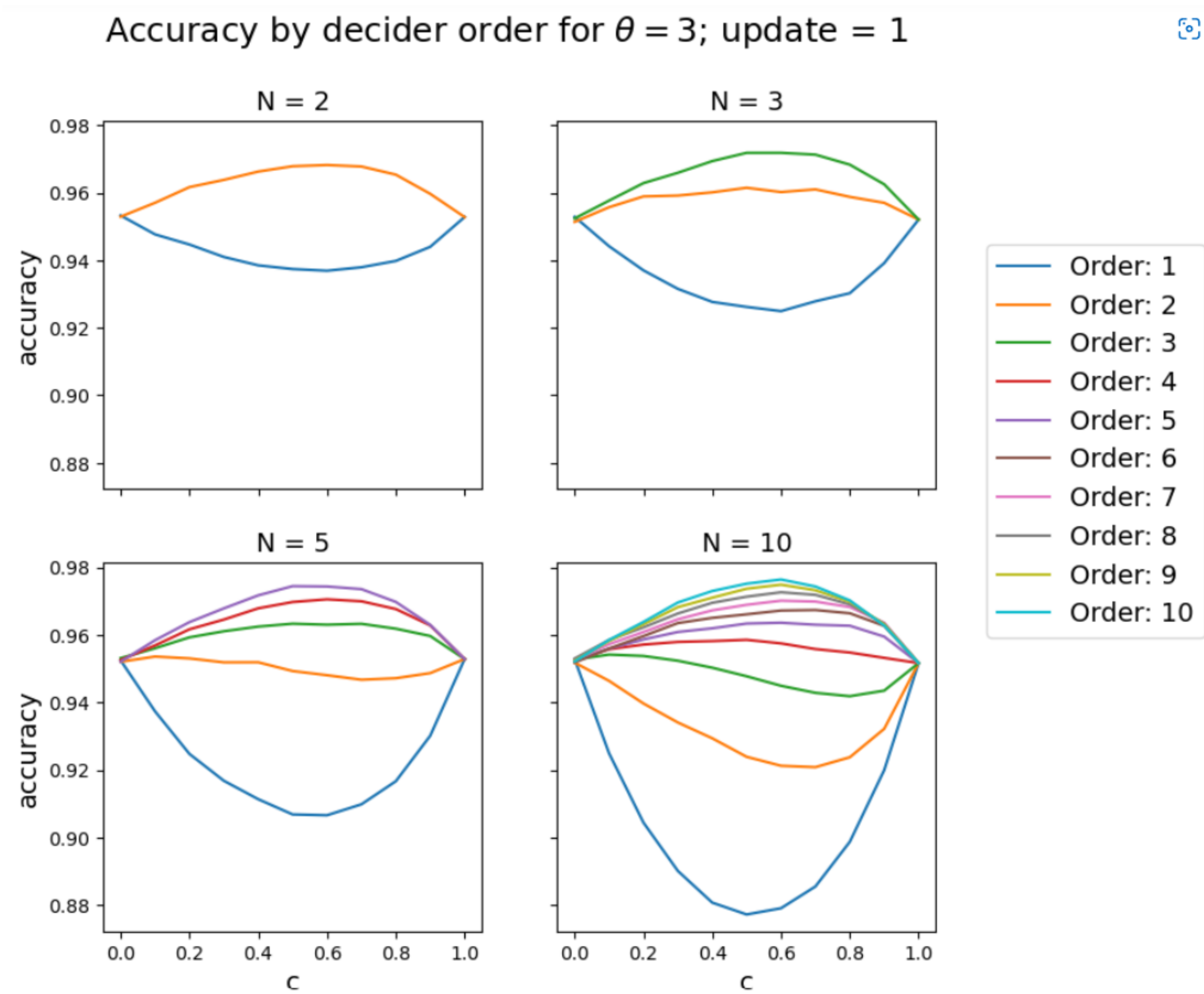

Figure 7.1: Accuracy by decider order when $T_{min} > 2$. Simulations show average decision accuracy for groups of agents each collecting private information until they reach threshold. Agents are numbered in the order in which they reach threshold. Accuracy is calculated by averaging the accuracy for the i^{th} decider over all trials. Accuracy increases as a function of decision order without respect to decision time. Simulations for threshold $\theta = 3$, update size $\upsilon = 1$.

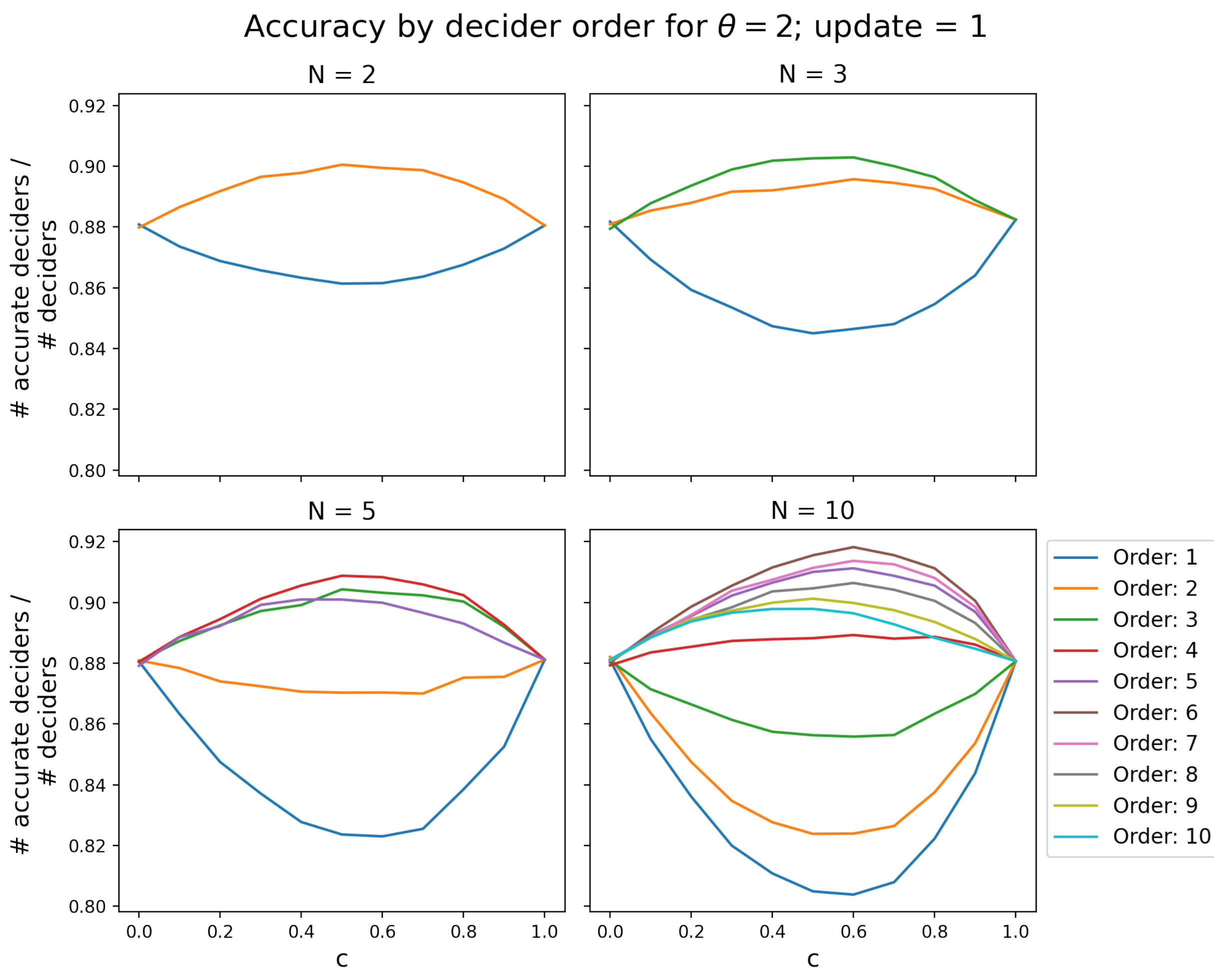

Figure 7.2: Accuracy by decider order when $T_{min} = 2$. Unlike when $T_{min} > 2$ (compare Figure 7.1), the most accurate decider is a middle decider with decider accuracy steadily decreasing as the decider order increases past a certain point. Simulation results for threshold $\theta = 2$, update size $\upsilon = 1$.

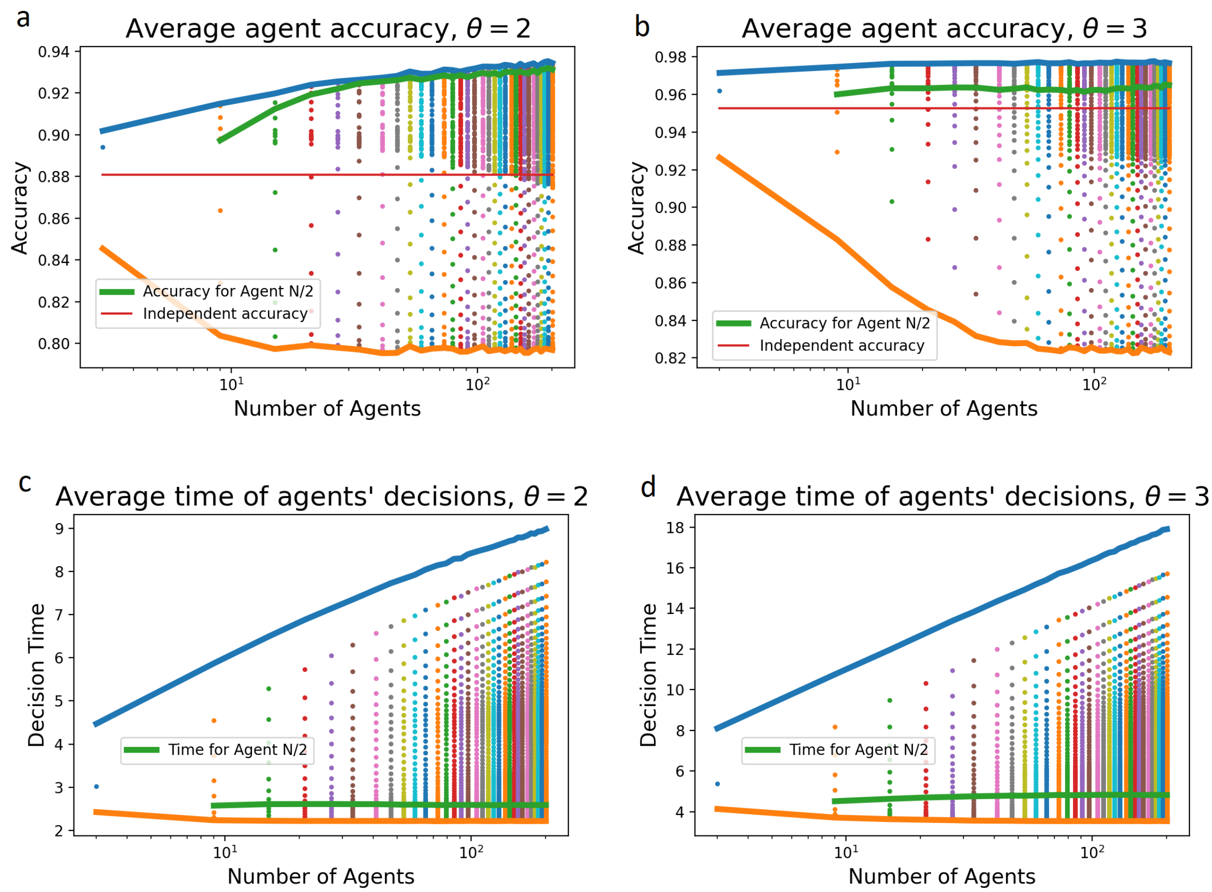

Figure 7.3: Average agent accuracy and average time of first decision for fixed probability $c = 0.5$ of receiving a common observation on each timestep and update size $\upsilon = 1$. Vertical dots of same color represent agents in a same-sized clique; higher dots correspond to later deciders. Averages are taken over the i^{th} decider in each trial. Panel a gives average agent accuracy for threshold $\theta = 2$. While the accuracy of the first decider quickly reaches a lower bound as N increases, the accuracy of the most accurate decider in the clique steadily increases with N. In panel b, which gives average agent accuracy for threshold $\theta = 3$, we see that while the accuracy of the first decider quickly reaches a lower bound, the accuracy of the most accurate decider even more quickly reaches an upper bound. In panels c and d, we see that while the time required for the last decider to reach threshold increases exponentially, the decision time of the first decider rapidly reaches the minimum possible time to decision (right). For both $\theta = 2$ and $\theta = 3$, the decision time of the middle decider plateaus just a little above the minimum possible time.

the N^{th} decider, suggesting that one effect of the presence of common (correlated) evidence in a non-communicating population might be to enhance the performance of intermediate deciders.

After acquiring a useable formula for first decider accuracy in the continuous correlated-evidence case, we will be well-positioned to commence a study on a correlated-evidence version of the social decision making model we examined in Chapters 2-4.

Appendix A

Notations

Notations given by the section in which they first appear. Notations whose usage is limited to the section in which they first appear are tagged *local only*. Notations whose usage is widespread are tagged **Recurring.** Notations which are similar to notations appearing in other sections have notes intended to DISAMBIGUATE by comparing the given notation to notation appearing in other sections.

· $\mathbb{E}[]$ - expectation of a random variable

· $\mathbb{V}[]$ - variance of a random variable

Section 2.1:

· H^0, H^1 : alternative true environmental states. *local only.*

· H_0, H_1 : hypotheses which are satisfied when H^0, H^1 are the true environmental states. *local only.*

· $p_0(x)$, $p_1(x)$: probability distributions of observations corresponding to true environmental states H^0, H^1. *local only.*

· $H^{\pm}$: alternative true *environmental states*. **Recurring.**

· $p_+(x)$, $p_-(x)$: probability distributions corresponding to environmental states H^+, H^-. The bulk of $p_+(x)$ occurs for $x > 0$, the bulk of $p_-(x)$ for $x < 0$. $p_+(x) = -p_-(x)$. **Recurring.**

· $P^+(\cdot)$: = $P(\cdot|H^+)$; the probability of $(\cdot)$ conditioned on H^+ being the true environmental state. **Recurring.**

· $P^-(\cdot)$: = $P(\cdot|H^-)$; the probability of $(\cdot)$ conditioned on H^- being the true environmental state. Non-recurring.

· A, B : positive and negative thresholds for Wald's likelihood update for the SPRT. Notation taken from [108]. *local only.*

· $\xi = \{x_1, x_2 ... x_t ...\}$: the trajectory ξ is an infinite sequence of observations drawn from $p_{\pm}$ when $H^{\pm}$ is the true environmental state. *local only.* DISAMBIGUATE : $\xi = (\mathbf{y}(t), x(t))$ in Section 6.1.

· $T(\xi)$: *Decision time*. The minimal time at which the belief of an agent reaches the upper or lower threshold. In a group with more than one agent, *first decision time* - the minimal time at which any agent reaches their upper or lower threshold. See T, T_i in Section 2.4. **Recurring.**

· $\text{LLR}(x)$, $\text{LLR}(\xi_{1:t})$: $= \log \frac{P^+(x)}{P^-(x)}$, $\log \frac{P^+(\xi_{1:t})}{P^-(\xi_{1:t})}$. *Log-likelihood ratio*; used to compute belief based on environmental observations. **Recurring.**

· $y(t)$, $y_i(t)$: $= \text{LLR}(\xi_{1:t})$. The *belief* (accumulated evidence) of an agent (agent i) at time t. If working with a continuous model, $y(t) = \int_0^t \frac{dy}{dt}$ (see Section 2.2). **Recurring.**

· θ, θ_i : magnitude of symmetric *threshold*, magnitude of symmetric threshold of agent i. A decision is triggered when an agent's belief (accumulated evidence) reaches θ or $-\theta$ (θ_i or $-\theta_i$). **Recurring.**

Section 2.2:

· α : coefficient on the drift term in the stochastic differential equation $dy = \alpha dt + \sqrt{2}dW$. $\alpha = \pm 1$ corresponding to environmental state $H^{\pm}$. *local only.* DISAMBIGUATE: $\alpha_{i,W}$ in Section 2.5, α in Section 5.3.

· W : standard Weiner process. **Recurring.** DISAMBIGUATE: W in Section 2.5, W_j in Section 6.4, $W_{\hat{i}}$ in Section 6.5.

· $dy := dt + \sqrt{2}dW$. Evidence accumulation SDE for private evidence in continuous model used in Chapters 2-4. **Recurring.**

Section 2.4:

· N : number of agents in a group of agents. **Recurring.**

· $d(y_i(t))$: the *decision state* of agent i at time t. $d(y_i(t)) = \pm 1$ if agent i has made a decision for $H^{\pm}$ and is 0 if agent i is still undecided.

· T_i : the time at which agent i makes a decision.

· $T := \min_{1 \leq i \leq N} T_i$. *First decision time*. See $T(\xi)$ in Section 2.1.

· $y_{priv}^{(i)}(t)$, $y_{soc}^{(i)}(t)$: the amounts of private and social evidence agent i has accumulated at time t. In the model for Chapters 2-4, $y_i(t) = y_{priv}^{(i)}(t) + y_{soc}^{(i)}(t)$ and $y_{priv}^{(i)}(t) = \int_0^t (dt + \sqrt{2}dw)$. For $y_{soc}^{(i)}(t)$, see $Soc(t)$ in Section 2.4.

· a, b : boundaries on the interval in which an agent's private evidence lies.

· $Soc_i(t) := \text{LLR}(y^{(i)}_{priv}(t) \in (a,b))$. The amount of evidence some agent $j \neq i$ can gain by observing the decision state of agent i at time t.

· $Soc(t) := \sum_{1 \le j \le N} Soc_j(t)$. The total amount of social evidence available. ($y^{(i)}_{soc}(t) = Soc(t) - Soc_i(t) = \sum_{j=1, j \neq i}^{N} Soc_j(t)$).

· $p^*_{\pm}(x,t)$: the distribution of agents' beliefs evolving according to the SDE $dy = dt + \sqrt{2}dW$, conditioned on no agent's belief having left the interval $(-\theta, \theta)$ at any time previous to t. **Recurring.**

Section 2.5:

· A_W : the set of agents reaching a decision in wave number W.

· W : wave number. **Recurring.** DISAMBIGUATE: W in Section 2.2, W_j in Section 6.4, $W_{\hat{i}}$ in Section 6.5.

· T^W : the time at which decisions in wave number W occur. *local only.*

· $\alpha_{i,W}$, $\beta_{i,W}$: prospective boundaries on the interval in which the private evidence of agent i lies based solely on the information made available in wave number W. *local only.* DISAMBIGUATE: α in Section 2.2, α in Section 5.3.

· $a_{i,W}$, $b_{i,W} := \max_{1 \le m \le W}(\alpha_{i,m})$, $\max_{1 \le m \le W}(\beta_{i,m})$. Boundaries on the interval in which the private evidence of agent i lies based on all information available at time T^W. *local only.* DISAMBIGUATE: a_W, a_1 in Section 3.2, a_i in Section 6.2, a_{R_j} in Section 6.3.

Section 3.1:

· $u(x,t)$, $u(x,t;x_0)$: solutions to a Smoluchowski equation with an initial point mass at 0, x_0. *local only.*

Section 3.2:

· $R_+(T)$: amount of social information made available by a single agent following decisions in the first wave following a first decision made at time T. **Recurring.**

· a_W, a_1 : number of agents in wave number W, number of agents in first wave (*size of first wave*). **Recurring.** DISAMBIGUATE: $a_{i,W}$ in Section 2.5, a_i in Section 6.2, a_{R_j} in Section 6.3.

· u_W : number of agents still undecided after wave number W.

· $c_1^{\pm}$: social increment $y_{soc}(T^1)$ to each agent after the first wave following a first decision for $H^{\pm}$. **Recurring.** DISAMBIGUATE: $\hat{c}_1^{\pm}$ in Section 3.6, c in Section 5.1.

Section 3.3:

· $\rho_{\pm}(x)$: *first passage time* distributions for a single agent through absorbing boundaries at $\pm\theta$.

· $\Phi_{\pm}(x,t)$: survival probabilities for passage of a single agent through absorbing boundaries at $\pm\theta$. $\Phi_{\pm}(x,t) = \int_0^t \rho_{\pm}(s)ds$.

· τ, τ_N : first passage time through boundaries $\pm\theta$ for a single agent; first passage time through boundaries $\pm\theta$ for the first agent in a group of N agents.

· $p_N(t)$: distribution for first decision times in a group of N agents. DISAMBIGUATE: $p_A(t)$ in Section 3.7, $p_{N,d}(t)$ in Section 4.1, $p_{k,+}$ in Section 4.10.

Section 3.6:

· $\hat{c}_1^{\pm}$: expected size of increment $c_1^{\pm}$ after the first wave. **Recurring.**

Section 3.7:

· $p_A(t)$: the probability a long agent undecided at time t has private belief that satisfies $y_{priv}^{(i)}(t) \geq 0$. *local only.* DISAMBIGUATE: $p_N(t)$ in Section 3.3, $p_{N,d}(t)$ in Section 4.1, $p_{k,+}$ in Section 4.10.

Chapter 4 introduction:

· θ_{min}, θ_{max} : smaller and larger threshold values in a population with dichotomous thresholds. (Each member of the population has either $\theta_i = \theta_{min}$ or $\theta_i = \theta_{max}$). **Recurring.**

· γ : fraction of the population with the lower threshold θ_{min}. **Recurring.**

Section 4.1:

· $p_{N,d}(t)$: first passage time distribution for N agents whose thresholds follow a dichotomous distribution. *local only*. DISAMBIGUATE: $p_N(t)$ in Section 3.3, $p_A(t)$ in Section 3.7, $p_{k,+}$ in Section 4.10.

· $P_{k,\pm}$: probability that the first decider has threshold k and decides correctly $(+)$ or incorrectly $(-)$.

· P_{min}, P_{max} : probability that the first decider has threshold θ_{min}, θ_{max}. **Recurring.**

Section 4.3:

· $R_{\pm,k}(T)$: information made available at the end of the fist wave cycle by an agent with threshold θ_k following a first decision for $H^{\pm}$. DISAMBIGUATE: R_+ in Section 3.2, R_j in Section 6.3, $R_{+,j}(\theta, y_j(\theta))$ in Section 6.4, $\mathbf{R}_{+,\hat{i}}$ in Section 6.5.

Section 4.5:

· $\Delta\theta := \theta_{max} - \theta_{min}$. Distance between largest and smallest thresholds. **Recurring**.

Section 4.10:

· $p_{k,+}$: probability that the first decider has threshold θ_k given that the first decision is correct. *local only.* DISAMBIGUATE: $p_N(t)$ in Section 3.3, $p_A(t)$ in Section 3.7, $p_{N,d}(t)$ in Section 4.1.

Section 5.1:

· c, $1-c$: the probabilities that on a particular timestep agents will all receive the common observation (c) or individual observations $(1-c)$. **Recurring.**

· $\eta_{\pm}$: observations favoring $H^{\pm}$. *local only.*

· p, q : probabilities of receiving an observation that favors (p) or disfavors (q) the true environmental state. **Recurring.**

Section 5.3:

· α : the probability that each agent makes an observation from the common pool or an individual observation on each timestep. *local only.* DISAMBIGUATION: α in Section 2.2, $\alpha_{i,W}$ in Section 2.5.

Chapter 6 introduction

· $y_{FD}(t)$: belief of the agent who is the first decider at time t.

Section 6.1:

· $\mathbf{y}(t)$: vector of beliefs of agents at time t. **Recurring.**

· $x(t)$: sequence recording 0 on each timestep for which all observations are common and 1 for each timestep on which all observations are independent. **Recurring.**

· $\xi(t) := (\mathbf{y}(t), x(t))$. The *trajectory* of the system at time t. **Recurring.** DISAMBIGUATE: ξ in Section 2.1.

· D : *Decision set*. The set of agents simultaneously reaching threshold at the time of the first decision. **Recurring.**

· D_+, D_- : the subsets of D that reached the positive threshold $+\theta$ or the negative threshold $-\theta$ at the time of the first decision T. **Recurring.**

· $\tilde{T}$: a specific value of the random variable $T(\xi)$. **Recurring.**

· $D_{\tilde{T}}$: the decision set at time $\tilde{T}$. The set of agents reaching threshold at the time of the first decision when the first decision occurs at time $T = \tilde{T}$. $D_{\tilde{T}} = \emptyset$ when $T \neq \tilde{T}$.

· $D_{\tilde{T},+}$, $D_{\tilde{T},-}$: subsets of $D_{\tilde{T}}$ containing the agents reaching positive and negative threshold, respectively. Both are empty when $T \neq \tilde{T}$. **Recurring.**

Section 6.2:

· a_i : number of observations in favor of environment H^+. DISAMBIGUATE: a_{R_j} in Section 6.3.

- $k(x(1:t)$: the total number of timesteps with independent observations by time t. When t is understood, may be given as $k(x)$ or just k. **Recurring.**

Section 6.3

- R_j : sequence of independent observations made by agent j up to time $\tilde{T}$. Determined from y_j, x. DISAMBIGUATE: R_+ in Section 3.2, $R_{+,j}(\theta, y_j(\theta))$ in Section 6.4, $\mathbf{R}_{+,\hat{i}}$ in Section 6.5.

- A_j : some countable set of sequences of independent observations R_j. DISAMBIGUATE: A_W in Section 3.2.

- a_{R_j} : the number of independent observations in favor of H^+ made by agent j up to time $\tilde{T}$. DISAMBIGUATE: a_i in Section 6.2.

Section 6.4:

- $R_{+,j}(\theta, y_j(\theta)) : = \theta - y_j(\theta))$. The distance between the belief of agent j and the belief of the first decider (assumed to be $y_{FD}(T) = \theta$. Takes on even integer values r_+ between 0 and 2θ. DISAMBIGUATE: R_+ in Section 3.2, R_j in Section 6.3, $\mathbf{R}_{+,\hat{i}}$ in Section 6.5.

- r_+ : a specific value of $R_{+,j}$. $r_+ \in \{0, 2, ...2\theta\}$.

- W_j : a subset of $\{0, 2, ...2\theta\}$ used to restrict values r_+ of $R_{+,j}$ to some part of the its range.

Section 6.5:

- $\mathbf{R}_{+,\hat{i}} : = (R_{+,1}, ...R_{+,i-1}, R_{+,i+1}...R_{+,N})$. A vector of $R_{+,j}$ values for every agent except agent i. DISAMBIGUATE: R_+ in Section 3.2, R_j in Section 6.3, $R_{+,j}(\theta, y_j(\theta))$ in Section 6.4.

· $W_{\hat{\imath}}$: some subset of the possible values of $\mathbf{R}_{+,\hat{\imath}}$.

· $\mathbf{r}_{+,\hat{\imath}}$: specific value of $\mathbf{R}_{+,\hat{\imath}}$. DISAMBIGUATE: r_+ in Section 6.4.

Section 6.6:

· $FD(D)$: the first decider, randomly chosen with equal probability from set D. **Recurring.**

Section 6.8:

· $\tilde{\theta}(k)$: *pseudothreshold* for a trial with k timesteps with independent observations. Gives accuracy conditioned on k for large N. **Recurring.**

· $\upsilon := \log \frac{p}{q}$. Magnitude of updates. **Recurring.**

· $\bar{t}$: gives the number of timesteps considered to have a random ordering of independent and common observations when approximating $P(k|T)$ for large N.

Bibliography

[1] Acemoglu, D., Dahleh, M., Lobel, I., and Ozdaglar, A. Bayesian learning in social networks. *National Bureau of Economic Research Working Papers Series* (2008).

[2] Alevy, J., Haigh, M., and List, J. Information cascades: Evidence from a field experiment with financial market professionals. *The Journal of Finance 62* (2007).

[3] Arganda, S., Perez-Escudero, A., and Polavieja, G. A common rule for decision making in animal collectives across species. *PNAS 109*, 50 (2012).

[4] Ashby, F. G. A biased random walk model for two choice reaction times. *Journal of Mathematical Psychology 27* (1983).

[5] Bahrami, B., Olsen, K., Latham, P., Roepstorff, A., Rees, G., and Frith, C. Optimally interacting minds. *Science 329* (2010).

[6] Banerjee, A. V. A simple model of herd behavior. *Quarterly Journal of Economics 107* (1992).

[7] Barth, H., Kanwisher, N., and Spelke, E. The construction of large number representations in adults. *Cognition 86* (2003).

[8] Beck, J., Ji Ma, W., Pitkow, X., Latham, P., and Pouget, A. Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron 74* (2012).

[9] Ben-Yashar, R., and Paroush, J. A nonasymptotic Condorcet jury theorem. *Soc. Choice Welf. 17* (2000).

[10] Bisazza, A., Butterworth, B., Piffer, L., Bahrami, B., Petrazzini, M., and Agrillo, C. Collective enhancement of numerical acuity by meritocratic leadership in fish. *Scientific Reports 4* (2014).

[11] Bogacz, R., Brown, E., Holmes, P., and Cohen, J. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review 113*, 4 (2006), 700–765.

[12] Bogacz, R., Wagenmakers, E.-J., Forstmann, B., and Nieuwenhuis, S. The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences 33* (2010).

[13] Boyd, R., and Richerson, P. *Culture and the Evolutionary Process*. The University of Chicago Press, 1985.

[14] Brown, R. *Social Psychology: The Second Edition*. The Free Press, 1986.

[15] Busemeyer, J., and Diederich, A. Survey of decision field theory. *Mathematical Social Sciences 43* (2002).

[16] Busemeyer, J., and Townsend, J. Decision field theory: A dynamic cognitive approach to decision-making in an uncertain environment. *Psychological Review 100*, 3 (1993).

[17] Caginalp, R., and Doiron, B. Decision dynamics in groups with interacting members. *SIAM Journal of Applied Dynamical Systems 16*, 3 (2017).

[18] Castanon, S., Moran, R., Ding, J., Egner, T., Bang, D., and Summerfield, C. Human noise blindness drives suboptimal cognitive inference. *Nature Communications 10* (2019).

[19] Charness, G., Karni, E., and Levin, D. Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty 35* (2007).

[20] Charness, G., Karni, E., and Levin, D. On the conjunction fallacy in probability judgement: New experimental evidence regarding linda. *Games and Economic Behavior 68* (2010).

[21] Conradt, L. Models in animal collective decision-making: information uncertainty and conflicting preferences. *Interface focus 2* (2011).

[22] Conradt, L., and Roper, T. Consensus decision making in animals. *TRENDS in Ecology and Evolution 20*, 8 (2005).

[23] Conradt, L., and Roper, T. Democracy in animals: The evolution of shared group decisions. *Proceedings of the Royal Society B 274* (2007).

[24] Couzin, I. Collective cognition in animal groups. *Trends in Cognitive Sciences 13* (2009).

[25] Couzin, I., Ioannou, C., Demirel, G., Gross, T., Torney, C., Hartnett, A., Conradt, L., Levin, S., and Leonard, N. Uninformed individuals promote democratic consensus in animal groups. *Science 334*, 6062 (2011), 1578–1580.

[26] Couzin, I., Krause, J., Franks, N., and Levin, S. Effective leadership and decision-making in animal groups on the move. *Nature 433* (2005).

[27] Davis, J. Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples 1950-1990. *Organizational Behavior and Human Decision Processes 52* (1992).

[28] de Condorcet, M. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'Imprimerie Royale, Paris, 1785.

[29] DeGroot, M. Reaching a consensus. *Journal of the American Statistical Association 69*, 345 (1974).

[30] DeGroot, M. A conversation with George A. Barnard. *Statistical Science 3*, 2 (1988).

[31] Degroot, M. *Optimal Statistical Decisions*. John Wiley & Sons, Ltd, Hoboken, 2004.

[32] Denter, P., Dumav, M., and Ginzburg, B. Social connectivity, media bias, and correlation neglect. *The Economic Journal 131* (2021).

[33] Drugowitsch, J., Wyart, V., Devauchelle, A., and Koechlin, E. Computational precision of mental inference as critical source of human choice suboptimality. *Neuron 92* (2016).

[34] Dyer, J., Ioannou, C., Morrel, L., Croft, D., Couzin, I., Waters, D., and Krause, J. Consensus decision making in humans crowds. *Animal Behaviour 75* (2008).

[35] Dyer, J., Johansson, A., Helbing, D., Couzin, I., and Krause, J. Leadership, consensus decision making and collective behaviour in humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 364* (2009).

[36] Eckert, J., Call, J., Hermes, J., Herrmann, E., and Rackoczy, H. Statistical inferences in chimapnzees and humans follow weber's law. *Cognition 180* (2018).

[37] Edwards, W. The theory of decision making. *Psychological Bulletin 51*, 4 (1954).

[38] Enke, B., and Zimmerman, F. Correlation neglect in belief formation. *Review of Economic Studies 86* (2019).

[39] Fechner, G. *Elemente der Psychophysik*, vol. 2. Breitkopf and Hartel, Leipzig, 1889.

[40] Fetsch, C., Kiani, R., Newsome, W., and Shadlen, M. Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron 83* (2014).

[41] Fisher, R. A., and Tippet, L. Limiting forms of the frequency distribution of the largest or smallest members of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society 24* (1928), 180–190.

[42] Franks, N., Richardson, R., Stroeymeyt, N., Kirby, R., Amos, W., Hogan, P., Marshall, J., and Schlegal, T. Speed–cohesion trade-offs in collective decision making in ants and the concept of precision in animal behaviour. *Animal Behaviour 85* (2013), 1233–1244.

[43] Gale, D., and Kariv, S. Bayesian learning in social networks. *Games and Economic Behavior 45* (2003).

[44] GLAESER, E., AND SUNSTEIN, C. Extremism and social learning. *Journal of Legal Analysis 1*, 1 (2009).

[45] GOLD, J., AND SHADLEN, M. The infludence of behavioral contexxt on the representation of a perceptual decision in developing oculomotor commands. *Journal of Neuroscience 23*, 2 (2003).

[46] GOLD, J., AND SHADLEN, M. The neural basis of decision making. *Annual Review of Neuroscience 30* (2007), 535–574.

[47] GOMEZ, P., AND PEREA, M. Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *Quarterly Journal of Experimental Psychology 67* (2014).

[48] GOOD, I. Studies in the history of probability and statistics: Xxxxvi. A. M. Turing's statistical work in World War II. *Biometrika 66* (1979).

[49] GRANOVETTER, M. Threshold models of collective behavior. *American Journal of Sociology 83*, 6 (1978).

[50] GREBE, T., SCHMID, J., AND STIEHLER, A. Do individuals recognize cascade behavior of others? – an experimental study. *Journal of Economic Psychology 29* (2008).

[51] HALLER, S., BANG, D., BAHRAMI, B., AND LAU, J. Group decision-making is optimal in adolescence. *Nature 8* (2018).

[52] HARTNETT, A., SCHERTZER, E., LEVIN, S., AND COUZIN, I. heterogeneous preference and local nonlinearity in consensus decision making. *Physical Review Letters 116* (2016).

[53] HAUN, D., REKERS, Y., AND TOMASELLO, M. Majority-based transmission in chimpanzees and human children, but not organgutans. *Current Biology 22* (2012).

[54] HAZLA, J., JADBABAIE, A., MOSSEL, E., AND RAHIMIAN, M. Bayesian decision making in groups is hard. *Operations Research 69*, 2 (2021).

[55] HERZ, D., ZAVALA, B., BOGACZ, R., AND BROWN, P. Neural correlates of decision thresholds in human subthalamic nucleus. *Current Biology 26* (2016).

[56] HOLMES, O. Abrams v. United States, 1919.

[57] JOLLES, J., KING, A., AND KILLEN, S. The role of individual heterogeneity in collective animal behaviour. *Trends in Ecology & Evolution 35* (2020).

[58] KALLIR, I., AND SONSINO, D. The neglect of correlation in allocation decisions. *Southern Economic Journal 75*, 4 (2009).

[59] Karamched, B., Stickler, M., Ott, W., Lindner, B., Kilpatrick, Z., and Josic, K. Heterogeneity improves speed and accuracy in social networks. *Physical Review Letters 121* (2020).

[60] Karamched, B., Stolarczyk, S., Kilpatrick, Z., and Josic, K. Bayesian evidence accumulation on social networks. *Siam J. Applied Dynamics Systems 19*, 3 (2020), 1884–1919.

[61] Kawamura, K., and Vlaseros, V. Expert information and majority decisions. *Journal of Public Economics 147* (2017).

[62] Khalvati, K., Park, S., Mirbagheri, S., Philippe, R., Sestito, M., Dreher, J.-C., and Rao, R. Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science Advances 5* (2019).

[63] Kimura, M., and Moehlis, J. Group decision-making models for sequential tasks. *SIAM Review 54*, 1 (2012).

[64] Kira, S., Yang, T., and Shadlen, M. A neural implementation of Wald's sequential probability ratio test. *Neuron 85* (2015).

[65] Krajbich, I., Hare, T., Bartling, B., Morishima, Y., and Fehr, E. A common mechanism underlying food choice and social decisions. *PLoS Computational Biology 11*, 10 (2015).

[66] Kugler, T., Kausel, E., and Kocher, M. Are groups more rational than individuals? a review of interactive decision making in groups. *WIREs Cognitive Science 3* (2012).

[67] Laming, D. Autocorrelation of choice-reaction times. *Acta Psychologica 43* (1979).

[68] Levy, G., and Razin, R. Correlation neglect, voting behavior, and information aggregation. *American Economic Review 105* (2015).

[69] Middleton, E., Reid, C., Mann, R., and Latty, T. Social and private information influence the decision making of australian meat ants (*Iridomyrmex purpureus*). *Insectes Sociaux 65* (2018).

[70] Milgrom, P., and Stokey, N. Information, trade, and common knowledge. *Journal of Economic Theory 26* (1982).

[71] Miller, N., Garnier, S., Hartnett, A., and Couzin, I. Both information and social cohesion determine collective decisions in animal groups. *PNAS 110*, 13 (2013).

[72] Morgan, T., Rendell, L., Ehn, M., Hoppitt, W., and Laland, K. The evolutionary basis of human social learning. *Proceedings of the Royal Society B 279* (2012).

[73] Moser, J., and Wallmeier, N. Correlation neglect in voting decisions: An experiment. *Economics Letters 198* (2021).

[74] Mossel, E., Sly, A., and Tamuz, O. Asymptotic learning on bayesian social networks. *Probability Theory and Related Fields 158* (2014).

[75] Moustafa, A., Keri, S., Somlai, Z., Balsdon, T., Frydecka, D., Misiak, B., and White, C. Drift diffusion model of reward and punishment learning in schizophrenia: Modeling and experimental data. *Behavioural Brain Research 291* (2015).

[76] Mueller-Frank, M. A general framework for rational learning in social networks. *Theoretical Economics 8* (2013).

[77] Mulder, M., van Maanen, L., and Forstmann, B. Perceptual decision neurosciences – a model-based review. *Neuroscience 277* (2014).

[78] Park, S., Goiame, S., O'Connor, D., and Dreher, J. Integration of individual and social information for decision-making in groups of different sizes. *PLoS Biology 15*, 6 (2017).

[79] Perez-Escudero, A., and Polavieja, G. Collective animal behavior from Bayesian estimation and proability matching. *PLoS Computational Biology 7* (2011).

[80] Perreault, C., Moya, C., and Boyd, R. A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior 33* (2012).

[81] Piantadosi, S., and Cantlon, J. True numerical cognition in the wild. *Psychological Science 28* (2017).

[82] Pillot, M.-H., Gautrais, J., Arrufat, P., Couzin, I., Bon, R., and Deneubourg, J.-L. Scalable rules for coherent group motion in a gregarious vertebrate. *PLoS ONE 6* (2011).

[83] R., C. D., and H.D, M. *The Theory of Stochastic Processes*. Methuen and Co. Ltd, 1965.

[84] Ratcliff, R. A theory of memory retrieval. *Psychological Review 85*, 2 (1978).

[85] Ratcliff, R., and McKoon, G. The diffusion decision model: theory and data for the two-choice decision task. *Neural Computations 20*, 4 (2008).

[86] Ratcliff, R., and Rouder, J. Modeling response times for two-choice decisions. *Psychological Science 9* (1998).

[87] Ratcliff, R., Smith, P., Brown, S., and McKoon, G. Diffusion decision model: current issues and history. *Trends in Cognitive Science 20*, 4 (2016).

[88] Rivers, D. Heterogeneity in models of electoral choice. *American Journal of Political Science* (1988).

[89] Saaty, T. *The analytic hierarchy process: planning, priority setting, and resource allocation*. McGraw-Hill International Book Co., New York, 1980.

[90] Schkade, D., Sunstein, C., and Kahneman, D. Deliberating about dollars: severity shift. *Columbia Law Review 100* (2000).

[91] Seeley, T., Camazine, S., and Sneyd, J. Collective decision-making in honey bees: How colonies choose among nectar sources. *Behavioral Ecology and Sociobiology 28* (1991).

[92] Shadlen, M., and Newsome, W. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology 4* (1994).

[93] Shadlen, M., and Newsome, W. Motion perception: Seeing and deciding. *PNAS 93* (1996).

[94] Smith, P. Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology 44* (2000).

[95] Smith, P., and Ratcliff, R. Psychology and neurobiology of simple decisions. *TRENDS in Neurosciences 27*, 3 (2004).

[96] Sorkin, R., Hays, C., and West, R. Signal-detection analysis of group decision making. *Psychological Review 108* (2001).

[97] Spiegler, R. Bayesian networks and boundedly rational expectations. *Quarterly Journal of Economics* (2016).

[98] Srivastava, V., and Leonard, N. Collective decision-making in ideal networks: The speed-accuracy tradeoff. *Transactions on Control of Network Systems 1*, 1 (2014).

[99] Stone, M. Models for choice-reactions time. *Psychometrika 25*, 3 (1960).

[100] Strandburg-Peshkin, A., Farine, D., Couzin, I., and Crofoot, M. Shared decision-making drives collective movement in wild baboons. *Science 348* (2015).

[101] Stubager, R., Seeberg, H., and So, F. One size doesn't fit all: Voter decision criteria heterogeneity and vote choice. *Electoral Studies 52* (2018).

[102] Sumpter, D., Krause, J., James, R., Couzin, I., and Ward, A. Consensus decision making by fish. *Current Biology 18* (2008).

[103] Tremel, J., and Wheeler, M. Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage 109* (2015).

[104] Turner, J., Hogg, M., Oakes, P., Reicher, S., and Wetherell, M. *Rediscovering the social group: A self-categorization theory*. Blackwell, 1987.

[105] Usher, M., and McClelland, J. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review 108*, 3 (2001).

[106] VICKERS, D. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics 13* (1970).

[107] VOSS, A., ROTHERMUND, K., AND VOSS, J. Interpreting the parameters of the diffusion model: An empirical validation. *Memory and Cognition 32*, 7 (2004).

[108] WALD, A. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics 16* (1945), 117–186.

[109] WALD, A., AND WOLFOWITZ, J. Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics 19* (1948), 326–339.

[110] WARD, A., HERBERT-READ, J., SUMPTER, D., AND KRAUSE, J. Fast and accurate decisions through collective vigilance in fish shoals. *PNAS 108*, 6 (2011).

[111] WARD, A., SUMPTER, D., COUZIN, I., AND KRAUSE, J. Quorum decision-making facilitates information transfer in fish shoals. *PNAS 105*, 19 (2008).

[112] WATTS, D. A simple model of global cascades on random networks. *PNAS 99*, 9 (2002).

[113] YANIV, I. The benefit of additional opinions. *Current Directions in Psychological Science 13* (2004).