The Wisdom of the Minority

Steven Callander Johannes Hörner*

September 28, 2005

Abstract

That later agents follow the herd of earlier actions is a central insight of research into economic cascades and social learning. We consider a variant of the standard model in which agents are differentially informed and observe only the market shares of the competing alternatives, and not the entire decision sequence. In this environment we show, for a large range of parameter values, that following the herd is not equilibrium behavior and that agents are better served following the minority, even if they themselves possess no private information and the minority consists of a lone dissenter against an arbitrarily large majority. In such cases, therefore, the minority is wiser than the majority.

1 Introduction

Consider the following decision problem.[1] A tourist is faced with the choice between two restaurants about which he has no private information. He can observe – by peering through the restaurant windows – that three patrons are in one restaurant and only one in the other. In the town there is a mix of tourists and locals, the locals know the relative quality of the restaurants but otherwise are indistinguishable from tourists. If the diners arrived at the restaurants in random order one at a time (facing a similar decision problem), which restaurant should the tourist choose?

In this paper we show that *following the herd* is often the wrong answer. We show that, after any history, following the (nonempty) minority is the optimal choice, provided the fraction of informed locals is not too great; for instance, if this fraction does not exceed 43% and the minority consists of a lone dissenter, then following the minority is *always* the optimal strategy, no matter how large the majority. A minority, therefore, can be wiser than the majority.

*Kellogg School of Management, Northwestern University, Evanston, IL 60208. For helpful comments we thank David McAdams, Meg Meyer, and Peter Norman Sørensen.

[1]A version of this example first appeared in Banerjee (1992); see also Chamley (2004).

The model we develop is a variant of the canonical model of social learning introduced by Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992), hereafter BHW. Our model differs from the canonical model in two subtle but significant respects that, we argue, capture more accurately the underlying economic environment. We assume that agents are differentially informed (tourists versus locals) and, most notably, that they observe only the number of patrons in each restaurant.

In contrast, in the model of Banerjee and BHW it is assumed that agents observe both the number of diners in each restaurant *and* the order in which they arrived. The canonical model is equivalent to assuming that the hypothetical tourist can peer through the windows and not only count the patrons inside but also note whether they are currently enjoying their appetizers, mains, or desserts (and from this information infer the order of arrival).

We show that these modelling differences are important. An optimizing agent still attempts to infer the private information of the seated patrons, but this inference is now confounded by uncertainty over the order in which they arrived. Did the lone dissenter arrive first? Or did he arrive last? The answer to this question is critical to any inferences drawn from the sequence of actions. If the dissenter arrived first then one may believe that he is a tourist who was subsequently not followed by a local, whereas if he arrived last then it would be reasonable to believe that he is himself a local, deviating from the inferior choices of tourists.

Simple calculations are able to establish that minority wisdom is possible in general environments. Tractability, however, limits our formal statements to a simple model that is little more than a formalization of the above example. In the model there are two states of the world and agents are either fully informed about the preferred state (they are locals) or entirely uninformed (they are tourists). All agents possess common preferences over consuming the better product irrespective of the state of the world. This simple model captures the possibility of minority choice in its starkest form and possesses the additional advantage of focusing attention on social learning, avoiding the interplay of public and private information (as informed agents always choose correctly and uninformed agents have no private information).

In the simple environment just described we show that minority wisdom is pervasive: indeed, the strategy of following the majority in all circumstances is optimal (if and) only if the fraction of informed agents is greater than $\frac{7}{9}$. If instead informed agents compose less than $\frac{7}{9}$ of the population then, in some circumstances, agents find it optimal to follow the minority. More surprisingly, if the fraction of informed agents is small enough then, for any given minority size, following such a minority or a smaller one is optimal *no matter* how large the majority. In particular, if the population is finite, then all agents find it optimal to always follow the (nonempty) minority.

Closely related to individual behavior is the issue of aggregate efficiency. The standard conclusion from the cascade literature is that the formation of herds inhibits social learning. One may wonder, therefore, if the contrarian behavior characterized here enhances social learning. Unfortunately, although minority choice can break incorrect herds, the limited information available to subsequent agents makes it difficult for them to learn sufficiently from previous actions.

Withholding information about sequence, therefore, does not prove a panacea to the difficulties of social learning in incomplete information environments.

Agents in our model follow the minority if and only if it is in their interests to do so. As informational externalities are ignored, it is therefore not surprising that the equilibrium strategy is not efficient for arbitrary populations. Despite this failure, limit efficiency holds and the fraction of agents making the correct decision approaches one as the population becomes large. Limit efficiency implies that minority choice ultimately breaks down and cannot support an equilibrium for every possible history (as otherwise minority choice would eternally pull market shares towards $\frac{1}{2}$). Despite this, the equilibrium strategy, even after many agents have acted, remains complex and involves a mix of following the minority and majority.

Although our model is closely related to the restaurant example, its simplicity allows wider application of the key intuition. There are many decision environments in which agents observe only some form of aggregate statistic and not the full sequence of choices. Often in product markets the only available information is market shares. Similarly, stock prices are cumulative statistics of all past actions and are generally considered to be sufficient statistics for market beliefs. The possibility that sequence matters – as shown here for the canonical model of social learning – carries, therefore, broad implications for many environments and warrants closer inspection.

The inference problem at the heart of our analysis is also relevant to more general environments. In our model decision makers must draw inferences about what other agents knew when making their decisions. This problem becomes only more important as assumptions on the decision sequence and the structure of communication are relaxed. For example, in unmediated group discussions, each individual must additionally draw inferences about who each other person has communicated with. These issues also arise when social learning is set in a network environment.

Ours is not the first paper to relax the full observability assumption of Banerjee and BHW. Several papers have assumed that agents observe only a partial history: Smith and Sørensen (1997, 1999) suppose agents observe a random sample from the history of actions and cover a variety of informational environments, including the one studied here (which they refer to as the aggregates model), and Çelen and Kariv (2002) assume a number of most recent actions are observed. The emphasis in all of these papers is not on individual behavior as is the case here, but rather on the evolution of beliefs, aggregate outcomes, and the possibility that restricting information can improve social learning.

Similarly, the assumption that agents are differentially informed also appears in the literature. Smith and Sørensen (2000) analyze general signal structures and note what they refer to as the *overturning principle*, that one opposing signal (and action) can overturn many previous, less informative, signals. Beliefs after such an action are very much a product of decision sequence and are path dependent. Our results build on the overturning principle and show that, even absent information about the decision sequence (the path), following the minority is optimal.

Finally, our results on minority wisdom can be contrasted to the finding in Smith and Sørensen

(2001) that aggregate efficiency sometimes calls for *contrarian* behavior by individual agents. Smith and Sørensen's (2001) result is related to the need for optimal experimentation and is not a consequence of individual optimization as is the case here.

The remainder of the paper is organized as follows. The model is described in Section 2. Results are presented in Sections 3 and 4 according to whether a majority of agents are uninformed or informed, respectively. Section 5 considers issues of efficiency and in Section 6 we show via example the power of minority wisdom in broader environments. Section 7 concludes.

2 The Model

The model is a simple variant of the canonical model of social learning introduced by BHW. There are two equiprobable states of the world, $\Omega = \{\omega_0, \omega_1\}$ and a countable set of agents, $i = 1, 2, \ldots$. Each agent must take an action, either 0 or 1. The payoff of an agent is 1 if the state is ω_0 and he takes action 0, or if the state is ω_1 and he takes action 1, and 0 otherwise.

Agents move sequentially. Each agent is either *informed*, with probability p, or *uninformed*. These informational types are privately known and independently distributed. An uninformed agent assigns probability $\frac{1}{2}$ to either state. An informed agent learns the realized state. Alternatively, agents can be thought of as homogeneous but receiving signals of different accuracy (either uninformative with probability $(1-p)$ or perfectly informative with probability p). The parameter p is common knowledge.

Agent n observes how many agents, among the first $n-1$ agents, have taken action 1, but he does not observe the order of these decisions (Agent 1 observes nothing). Therefore, the information set of Agent $(m_0 + m_1 + 1)$ can be denoted by an ordered pair (m_0, m_1), where m_0 and m_1 are the number of agents having chosen action 0 and 1, respectively. It is clear that an informed agent always takes the correct action (i.e., 0 if ω_0, 1 otherwise). We therefore focus on the decision of uninformed agents. Let $\alpha(m_0, m_1)$ denote the probability with which an agent takes action 1 after history (m_0, m_1). (In an abuse of notation, we refer to *histories* as particular information sets observed by agents, and *paths* to describe actual sequences of choices.)

If indifferent (that is, if states are equally likely given (m_0, m_1)), we assume that an agent chooses either action with probability $\frac{1}{2}$. As results are sensitive to the tie-breaking rule, this issue is further discussed in Section 5. The decision problem of an agent is the following: given history (m_0, m_1), which of the two states is more likely? Because a definite tie-breaking rule has been adopted, the solution to this decision problem is unique. Therefore, we can solve for the unique sequential equilibrium of the game, starting with the decision problem of the first agent, and proceeding recursively. As the strategic aspects of this game are rather elementary, we will simply refer to the equilibrium strategy as the *optimal* strategy.[2]

[2]This does not imply that this strategy is socially optimal. Efficiency properties of the optimal strategy are discussed in Section 4.

3 Wise Minorities: $p < \frac{1}{2}$

3.1 An Illustration: Minorities of One

The possibility of minority wisdom – that the minority is more likely to have chosen correctly than the majority – arises independently of the size of the majority and minority for any given population, provided that the fraction of informed agents is sufficiently small. In this section we illustrate the logic behind this finding for minorities of the smallest possible size. Considering histories of the form $(1, m)$, we show that an uninformed agent is often better served following the lone dissenter rather than the overwhelming majority (the case of $(m, 1)$ is analogous).

To establish that following the minority is optimal for an uninformed agent at a particular history requires the calculation of all possible decision sequences, or paths, that can produce the history. To make these calculations several observations on equilibrium behavior are required. First, by the tie-breaking assumption and symmetry, an uninformed agent observing a history (m, m) is indifferent and, again by the tie-breaking assumption, he mixes equally over the two actions. More generally, the optimal strategy is "symmetric" around such a state (See Remark 1 in the appendix). Second, uninformed agents always follow a unanimous choice. That is, if all previous agents have chosen identically, such as in the history $(0, m)$, then an uninformed agent follows the unanimous choice. The uninformed agent can be sure no previous agent was informed that action 0 is optimal, but it may be possible that a previous agent was informed that action 1 is optimal. Although the evidence may be weak, it leans in favor of the unanimous choice and an uninformed agent optimally follows this choice.

3.1.1 The First Wise Minority: History $(1, 2)$

We are now in position to calculate the probability of either state given history $(1, m)$. The earliest possibility for a single agent to compose a minority is $m = 2$. Figure 1 depicts, for each possible state of the world, the possible paths to the history $(1, 2)$, where the history is represented by the corresponding Cartesian point. In state ω_1 (the second panel) there is only one possible path to history $(1, 2)$: agent 1 is uninformed and chooses action 0, agent 2 is informed and chooses action 1, and agent 3 is either informed or uninformed and chooses action 1. The probability of this path conditional on state ω_1 is given by

$$\begin{aligned}
\Pr\{1, 2 \mid \omega_1\} &= \left(\frac{1-p}{2}\right) p \left(p + \frac{1-p}{2}\right) \\
&= \left(\frac{1-p}{2}\right) p \left(\frac{1+p}{2}\right).
\end{aligned}$$

Observe that if agent 1 chooses action 1 then, conditional on state ω_1, history $(1, 2)$ is not possible. In fact, unanimous choice is guaranteed as it can be broken only by an informed agent observing an opposing signal, which is not possible in state ω_1. This fact proves critical to the

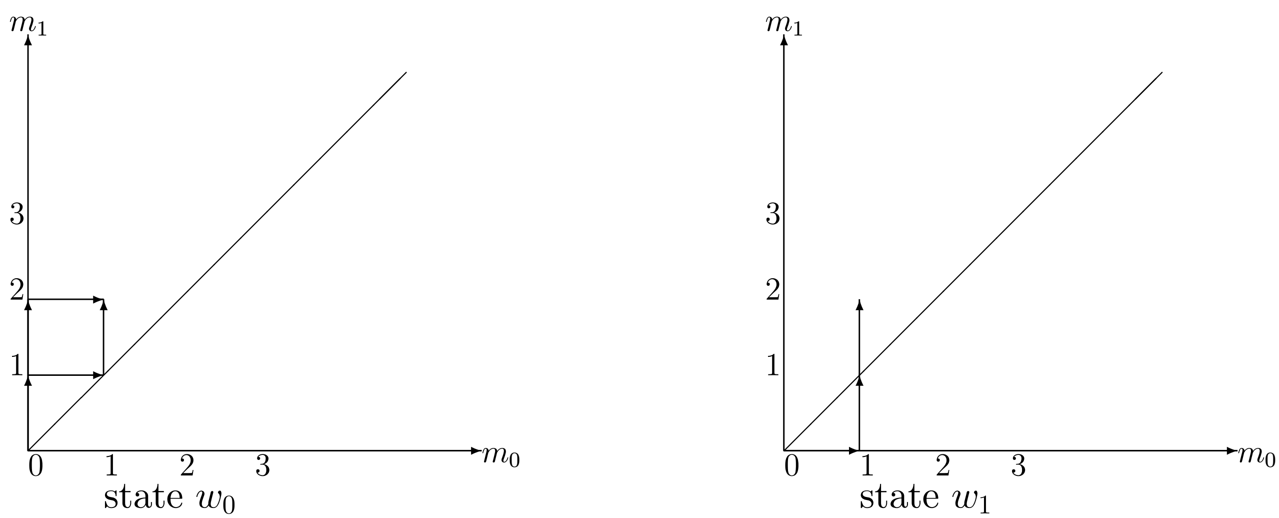

Figure 1: Paths to (1,2)

illustration presented here for minorities of size one, and similar ideas will prove critical to the general results of the following section.

For state ω_0, the history $(1,2)$ can be reached by two possible paths (the first panel of Figure 1), both require the first agent to be uninformed and to choose action 1. The paths diverge for agents 2 and 3 and are as follows: (i) agent 2 is informed and chooses 0, and agent 3 is uninformed and chooses action 1; (ii) agent 2 is uninformed and follows agent 1, and agent 3 is informed and chooses 0. The probability of these paths conditional on state ω_0 is

$$\begin{aligned}\Pr\{1,2 \mid \omega_0\} &= \left(\frac{1-p}{2}\right)\left(p\left(\frac{1-p}{2}\right)+(1-p)\,p\right)\\ &= \frac{3}{4}(1-p)^2\,p.\end{aligned}$$

An uninformed agent follows the minority if state ω_0 is more likely than state ω_1 given history $(1,2)$; that is, $\Pr\{\omega_0 \mid 1,2\} > \Pr\{\omega_1 \mid 1,2\}$. By Bayes' rule

$$\Pr\{\omega_0 \mid 1,2\} = \frac{\Pr\{\omega_0\}\Pr\{1,2 \mid \omega_0\}}{\Pr\{\omega_0\}\Pr\{1,2 \mid \omega_0\}+\Pr\{\omega_1\}\Pr\{1,2 \mid \omega_1\}}.$$

As the states are equally likely, the condition for minority choice reduces to

$$\Pr\{1,2 \mid \omega_0\} > \Pr\{1,2 \mid \omega_1\},$$

which by algebra requires

$$\frac{3}{4}(1-p)^2 p > \left(\frac{1-p}{2}\right) p \left(\frac{1+p}{2}\right)$$
$$\Leftrightarrow p < \frac{1}{2}.$$

Thus, if less than half of all agents are informed, an uninformed agent observing history $(1,2)$ optimally follows the minority and chooses action 0. This result is driven by two factors: the number of possible paths for each state, and the number of agents that are informed versus uninformed along each path. For history $(1,2)$ to be reached in state ω_0, one agent must be informed and two uninformed, and two possible paths reach this history. In contrast, for state ω_1 there exists only one possible path, but it is possible for either one or two agents to be informed. The more likely it is that an agent is informed, therefore, the more likely it is that state ω_1 produced the history. The value $p = \frac{1}{2}$ provides the cut-point on either side of which a different state is more likely to have produced the observed history.

3.1.2 A Lone Dissenter Versus Ever Larger Majorities

The finding of a wise individual at history $(1,2)$ is of interest in its own right, yet its true significance is revealed only when its impact on subsequent histories is considered. One may think that a wise individual is possible in a small group, but that ultimately an individual will be overwhelmed as the opposing majority grows. We show here that, in fact, the opposite is often true and that the relative wisdom of a lone dissenter can actually grow. This result is a product of the layering of decision problems as later agents not only face their own confounded information environment, but must allow for the confounding of previous agents' decision problems.

Suppose then that $p < \frac{1}{2}$ and at history $(1,2)$ an uninformed agent follows the minority. The calculations for histories $(1,m)$, where $m \geq 3$, are similar to those for $(1,2)$ although the logic of the result is slightly different. As depicted in Figure 2, only one path to $(1,3)$ exists for each possible state. Critical here is that should history $(1,2)$ be reached, then an uninformed agent follows the minority and induces history $(2,2)$. In state ω_0 (the first panel) the final agent must, therefore, have been informed and deviated from unanimous choice, as if the previous history had been $(1,2)$ the agent, irrespective of private information, would have chosen with the minority. For state ω_1 the history $(1,3)$ cannot be reached from history $(0,3)$ and the path can only come through history $(1,2)$ (and, therefore, the lone dissenter was the first agent). Consequently, for history $(1,3)$ to be reached in state ω_1 the 4th agent had to have been informed. The probabilities

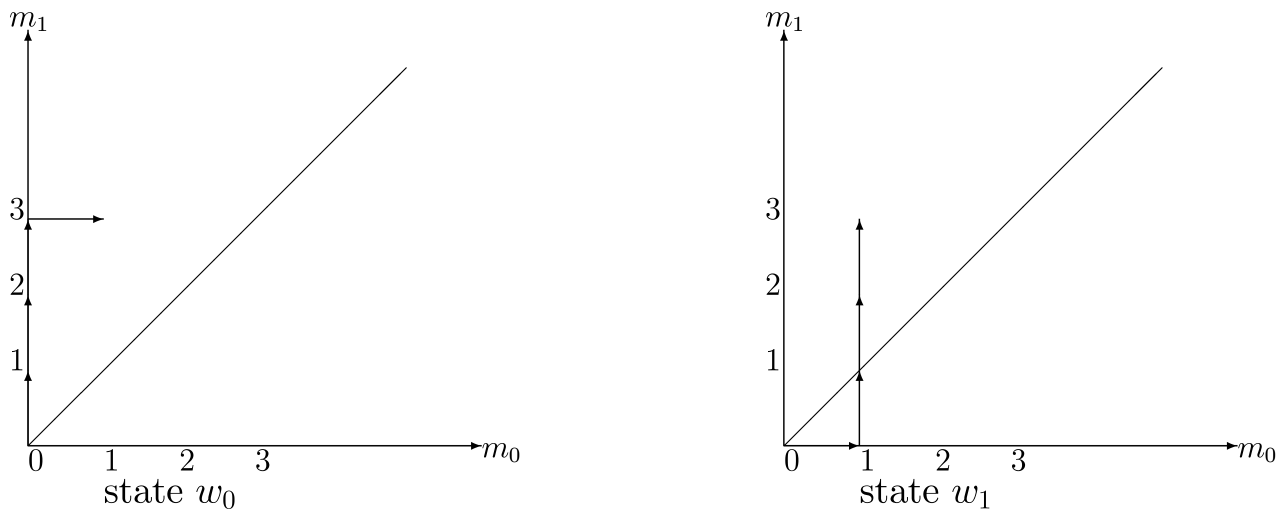

Figure 2: Paths to (1,3)

for these paths are as follows.

$$\begin{aligned}
\Pr\{1,3 \mid \omega_0\} &= \left(\frac{1-p}{2}\right)(1-p)^2 p && \text{(1a)}\\
&= \frac{1}{2}(1-p)^3 p. \\
\Pr\{1,3 \mid \omega_1\} &= \left(\frac{1-p}{2}\right) p \left(\frac{1+p}{2}\right) p && \text{(1b)}\\
&= \frac{1}{4}(1-p)(1+p)p^2.
\end{aligned}$$

Minority choice requires $\Pr\{\omega_0 \mid 1,3\} > \Pr\{\omega_1 \mid 1,3\}$, which by algebra, is equivalent to: $p < \frac{5-\sqrt{17}}{2} \approx 0.438$.

With an equal number of paths to reach $(1,3)$ for each state, the relative likelihood of the states hinges solely on the numbers of informed and uninformed agents. In state ω_0 there need only be one informed agent to reach history $(1,3)$, although this agent must be in a particular location. In contrast, for history ω_1 there is at least two informed agents – as due to minority wisdom at $(1,2)$, the 4th agent must be informed – and an additional agent (the third) has the luxury of being informed or uninformed. Thus, as informed agents are less prevalent than uninformed ones, minority wisdom decreases the probability of reaching $(1,3)$ in state ω_1. In fact, for a sufficiently small fraction of informed agents, it is more likely that one rather than two agents are informed and, therefore, more likely that the lone minority dissenter is informed rather than uninformed, and worthy of being followed.

The bound on minority wisdom for the history $(1,3)$ is tighter than for the history $(1,2)$, and following a minority may not be optimal at $(1,3)$ despite being optimal at $(1,2)$. Surprisingly perhaps, this contraction of the bound does not continue for larger populations when the minority is of size one. Therefore, as the wisdom of early minorities feeds into the decisions of later agents, a lone dissenter becomes increasingly attractive relative to a growing majority. Of course, if later agents knew that a lone dissenter had repeatedly opposed a growing majority (and the optimal strategy for earlier uninformed agents is to follow the minority), then their optimal decision would be obvious and very different to the current environment. What is important here is that later agents don't have this information. Thus, the prospect of a repeatedly ignored lone dissenter is overwhelmed by the probability that the dissenter was the most recent agent, and subsequent agents prefer to follow the apparently wise minority.

This result for ever increasing majorities – for histories $(1,m)$ – can be readily verified from the arguments above for history $(1,3)$. For $m>3$ there is again only one possible path for each history, and again the dissenter must be the final agent in state ω_0 and the first agent in state ω_1. Note, however, that as minority choice applies at $(1,2)$, $(1,3)$, and so on, there must be at least $(m-1)$ informed agents in state ω_1 but only one informed agent in state ω_0. The probabilities of reaching $(1,m)$, conditional on each state, are therefore generalizations of Equations (1a) and (1b) and are as follows.

$$\begin{aligned}\Pr\{1,m \mid \omega_0\} &= \frac{1}{2}(1-p)^m p \\ \Pr\{1,m \mid \omega_1\} &= \frac{1}{4}(1-p)(1+p)p^{m-1}.\end{aligned}$$

As $p<\frac{1}{2}$, $\Pr\{\omega_0 \mid 1,3\} > \Pr\{\omega_1 \mid 1,3\}$ implies $\Pr\{\omega_0 \mid 1,m\} > \Pr\{\omega_1 \mid 1,m\}$ and minority choice persists. Therefore, if an uninformed agent is prepared to follow a lone dissenter in a population of size four, it is optimal for an uninformed agent to follow a lone dissenter in populations of arbitrary size. In fact, the ratio of probabilities, $\Pr\{1,m \mid \omega_0\} / \Pr\{1,m \mid \omega_1\}$, is increasing in m and the relative wisdom of the minority is increasing in the size of the opposing majority.

3.2 Larger Minorities

Somewhat surprisingly, our analysis implies that not only may a lone dissenter be wiser than a majority, but that the dissenter's relative wisdom is increasing in the size of the opposing majority. In this section we show that wise minorities are not restricted to lone dissenters, but rather can arise for minorities of arbitrary size. The relative wisdom does, however, weaken as the minority grows and, ultimately, breaks down. Thus, for a fixed proportion of informed agents, a history must be reached such that following the majority is the optimal strategy.

We illustrate the rationale behind these possibilities for minorities of size three. The calculations are similar for all opposing majorities but, as with minorities of size one, a distinction exists between whether the majority exceeds the minority by one or more (history $(1,2)$ relative to histories $(1,m)$). We begin, therefore, with history $(3,4)$ and assume minority wisdom has obtained for all minorities up to three. Figure 3 describes the paths that lead to such a history under each state of the world.

The wisdom of the minority at history $(1,2)$ depended on both the number of paths under state ω_0 relative to state ω_1 and the number of informed versus uninformed agents. The same logic applies here for a larger minority, although the argument is weakened. This can be seen firstly be looking at the requirement for the first informed agent. Under either state, all paths must originate in an initial incorrect choice by an uninformed agent, followed shortly afterwards by the correct choice of an informed agent who breaks thereby the prevailing unanimity. If the correct state is ω_1 this informed agent must be one of the first four agents (obviously not the first), whereas if state ω_0 is correct then an informed agent must only be one of the first five. So, as in the case of a lone dissenter, there is one more opportunity here when the true state corresponds to the minority than otherwise. This pattern persists but the relative advantage diminishes: more generally, if the history is $(m_0, m_1) = (m, m+1)$, then, under ω_0, the probability of breaking from unanimity by the required points is $1-(1-p)^{m+1}$, while it is only $1-(1-p)^m$ under ω_1. The ratio of these two probabilities tends to $1+1/m$ as p tends to 0 (indeed, recall when $m=1$ that there are two paths leading to the history $(1,2)$ under ω_0, but only one under ω_1), which itself approaches 1 as m becomes large.

Moreover, further weakening the wisdom of the minority, the path after the first informed agent is not determinate (as is the case with a minority of 1) and, in fact, there are more paths leading to history $(3,4)$ under ω_1 than under ω_0. This occurs as history $(3,4)$ under state ω_1 is more consistent with longer strings of consecutive informed agents following a tie: for instance, from the history $(2,2)$, history $(3,4)$ can be reached in state ω_0 only through $(3,3)$ (passing through $(3,2)$ or $(2,3)$), while in state ω_1 there is an additional path to $(3,4)$ through history $(2,4)$. Of course, such a path involves one more informed agent, as only such an agent would choose action 1 at history $(2,3)$, but observe that this path is not the only one: moving backwards, and excluding the path just mentioned, there are two additional paths starting from $(2,2)$ (each involving two additional informed agents).

As the additional paths under state ω_1 require at least one additional informed agent, this effect, for fixed m, remains dominated by the advantage to state ω_0 (more flexibility in the location of the first informed agent), provided the proportion of informed agents is low enough. However, as m grows, the second effect ultimately dominates the first, and for fixed p the wisdom of the minority ultimately breaks down for some history $(m, m+1)$.

As with histories $(1,m)$, the logic for larger minorities is rather different when the majority exceeds the minority by more than one choice. Figure 4 describes the paths for history $(3,5)$ under both states. For state ω_0 there is now a unique path leading to $(3,5)$, along which a determinate agent, agent six, was informed and broke the prevailing, mistaken unanimity. Previous agents

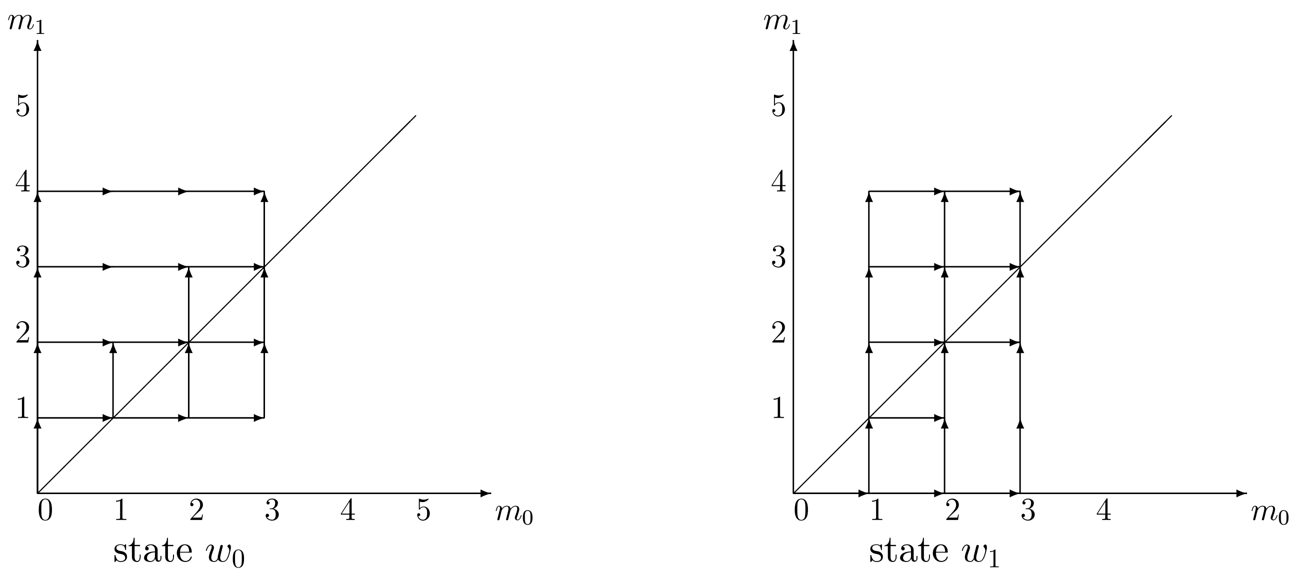

Figure 3: Paths to (3,4)

must all have been uninformed, and the type of later agents is irrelevant. Significant here is that this history is consistent with a single informed agent. In contrast, the pattern under state ω_1 is similar to the one that obtained for history $(3, 4)$: any given path requires at least two informed agents, but the number of paths is large and increasing in the size of the minority, as the first informed agent may have been one of many agents. Thus, in this case as well, minority wisdom prevails provided p is sufficiently close to zero, but breaks down for fixed p provided the minority size is large enough.

3.3 Minority Choice: Properties of the Optimal Strategy for $p < \frac{1}{2}$

This section formalizes and generalizes the arguments described so far. As we have seen, for any $p < \frac{5-\sqrt{17}}{2}$, minority choice is optimal independently of the majority size, provided only the existing minority does not exceed one agent. A similar result holds for larger minorities, although the required bound must be strengthened. In fact, the larger the required minority, the smaller is the critical bound on p such that minority wisdom obtains for that minority and all smaller minorities. If p_j denotes this bound for minorities of j or smaller, so that $p_1 = \frac{5-\sqrt{17}}{2}$, this means that p_j is strictly decreasing in j. In fact,

$$\lim_{j \to \infty} p_j = 0.$$

Therefore, there is no $p > 0$ for which minority choice would be optimal, independently of both the minority and the majority size. That is, for any $p > 0$, there always exists a history (m_0, m_1)

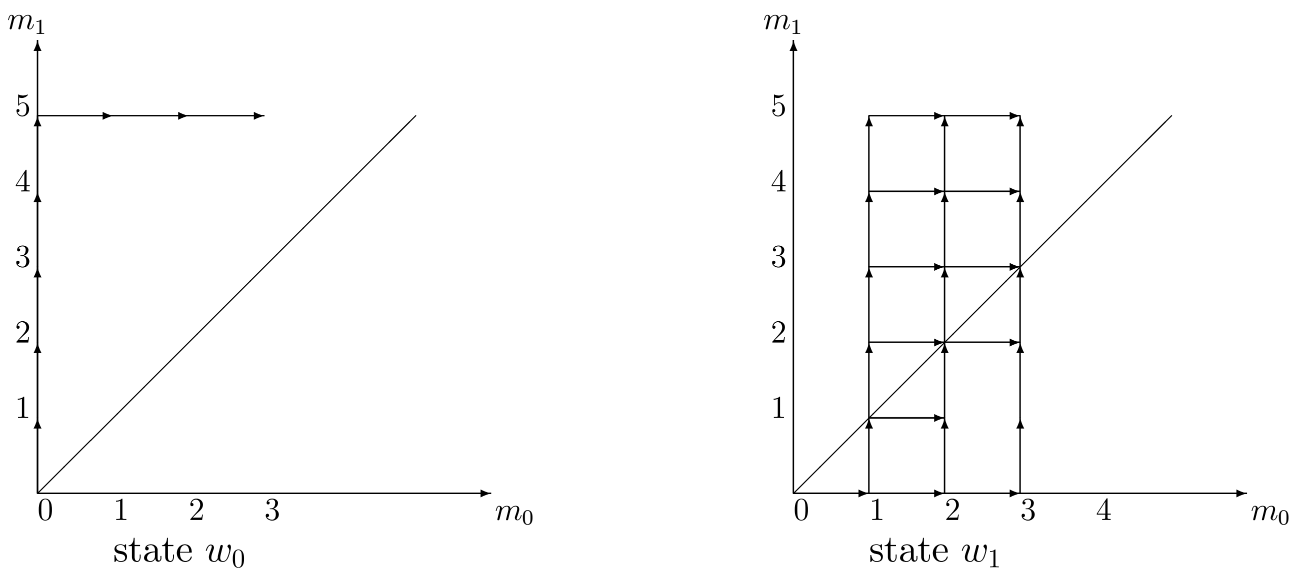

Figure 4: Paths to (3,5)

at which point it is optimal to follow the majority.[3] This discussion is summarized in the following Proposition:

Proposition 1 *(i) For any history, following the (nonempty) minority is optimal if the fraction of informed agents is smaller than some upper bound that is independent of the majority size:*

$$\forall j > 0, \exists p_j > 0, \forall p \in (0, p_j), \forall 0 < m_0 \leq j, \forall m_1 > m_0, \alpha(m_0, m_1) = 0.$$

(ii) There always exist histories after which choosing with the majority is optimal:

$$\forall p > 0, \exists (m_0, m_1), m_1 > m_0 > 0 \text{ such that } \alpha(m_0, m_1) = 1.$$

Proof: Part (i), which is a tedious generalization of the argument presented in the illustration, is relegated to the Appendix.

Part (ii): Suppose otherwise. That is, suppose that, for some $k > 0$, minority choice is optimal: $\forall m > k' \geq k$, $\alpha(m - k', m) = 0$. In this case, if the state of nature is ω_0, observe that:

$$\Pr\{m - k', m \mid \omega_0\} = \frac{1}{2}(1-p)^m p,$$

[3] In fact, infinitely many such histories exist. It is not true, however, that following the majority *thereafter* is necessarily optimal.

for the only path that connects $(0,0)$ to $(m-k',m)$, conditional on state ω_0 consists of a string of m uninformed agents who all chose action 1 (the first of which chose that action at random), immediately followed by one informed agent (the agents that then followed chose action 0 independently of their information).

Suppose now that the state of nature is ω_1. For $0 \leq j \leq m-k'$, consider the path along which the first j agents are uninformed, all of which choosing action 0, immediately followed by one informed agent, and then, after another arbitrary $\max\{j-k',0\}$ agents (so that the 'band of length k' around the diagonal' is reached), by as many uninformed agents as necessary to obtain that, overall, $m-k'$ agents have chosen action 0 (this cannot require more than another $m+k'-j$ agents). When this occurs, the number of agents having chosen action 1 must necessarily be between $m-2k'$ and m. Therefore, the probability of reaching $(m-k',m)$ following such a path is at least

$$\frac{1}{2}(1-p)^{m+k'}p^{2k'+1}.$$

Since there are $m-k'$ possible choices for the integer j, this means that:

$$\Pr\{m-k',m \mid \omega_1\} \geq \frac{1}{2}(m-k')(1-p)^{m+k'}p^{2k'+1}.$$

For fixed k', this number may be chosen to be arbitrarily large relative to $\frac{1}{2}(1-p)^m p$, by picking m large enough. This implies that, for such an m,

$$\Pr\{m-k',m \mid \omega_1\} > \Pr\{m-k',m \mid \omega_0\},$$

so that the optimal action after such a history is action 1, yielding the desired contradiction.■

It is easy to compute the first terms of the sequence of upper bounds $\{p_j\}$ recursively: $p_1 \simeq 0.44$, $p_2 \simeq 0.29$, $p_3 \simeq 0.17$, $p_4 \simeq 0.08, \ldots$, but a general formula appears elusive. In the proof of the first part of the proposition, the following lower bound is shown to hold: $p_j > 4^{-j}$. The breakdown of minority wisdom can arise for varying histories and does not depend on the minority being close in size to the opposing majority. Indeed, in the proof of the proposition it is shown that majority choice can arise even if the majority is arbitrarily larger than the minority; that is, for arbitrarily large $|m_1 - m_0|$.

It follows immediately from Proposition 1 that following the minority is optimal, independently of the minority size, provided that the population size does not exceed some N, as long as p is small enough. More formally, consider the following symmetric minority choice strategy α^*, defined by, for all agents $n \leq N$:

$$\alpha^*(m_0,m_1) = \begin{cases} 0 \text{ if } m_1 = 0, \\ 1 \text{ if } m_0 > m_1 > 1. \end{cases}$$

We have that, for $\bar{p} := p_{(n-2)/2}$ or $\bar{p} := p_{(n-1)/2}$ depending on the parity of n:

Corollary 1 *Given $N \in \mathbb{N}$, there exists $\bar{p} > 0$ such that, for all $p < \bar{p}$, α^* is the optimal strategy for all agents $n \leq N$.*

3.4 When Minority Wisdom Breaks Down

What can be said about the optimal strategy in general once minority wisdom breaks down? Unfortunately, a complete characterization does not appear possible. The optimal strategy has a complicated structure, prescribing majority choice in some circumstances and minority choice in others. The intricacies of the optimal strategy are illustrated by example in Figure 5 (where the arrow direction represents the optimal choice at a particular history and two arrows reflects indifference).

Despite the complexity, some approximate regularity does emerge. Figures 6 and 7 depict several features of choice when $p = 1/20$ and 100 agents have chosen. Figure 6 shows the optimal strategy for the 101st agent if uninformed (where the value represented is the probability of choosing action 1). As can be seen, minority choice has broken down (Corollary 1 no longer applies) and for two minority sizes it is optimal for an uninformed agent to follow the majority. This pattern (which appears generally in computations) shows that when the minority choice strategy α^* breaks down, the optimality of following the minority does not disappear altogether but remains rather pervasive.

Figure 7 depicts the distribution of market shares (represented by possible histories (m_0, m_1)) after 100 agents have chosen and when the state is ω_0. As the densities cover a wide range, the depiction has been replicated in two panels with different scales. The probability that $m_1 \leq 50$ is 0.765; therefore, after 100 agents have chosen, there is approximately a $\frac{1}{4}$ probability that the minority is wiser than the majority.

Although minority choice draws market shares towards $\frac{1}{2}$, several histories are reached with significant probability (the six points discernible from zero in the left panel). The boundary history, $m_1 = 0$ (corresponding to history $(100, 0)$), is easy to comprehend. The five internal histories are less easy to interpret, although they appear to indicate that the breakdown of minority choice is not inconsequential. Rather, the optimal strategy produces paths through the choice sequence that "collect" a significant fraction of histories. The second panel of Figure 7 reflects the pattern that one may expect from minority choice: that the probability of a particular history is greater the closer the market shares are to being equal. If the scale is contracted further then a similar pattern emerges for histories where m_1 is less than 50.

The distribution displays a fascinating pattern. "Jumps" in the density appear precisely at states where the optimal strategy switches from one action to another. That is, the three points appearing for minority sizes $m_1 < 50$ (including the case $m_1 = 0$), whose probability are significant enough to be discerned on the left panel are precisely the three values of m_1 for which it is optimal to follow the majority. The same observation applies to the case $m_1 > 50$. For those states $m_1 < 50$ for which following the minority is optimal, their probability increases geometrically as m_1 increases (roughly by a factor 10^3 from m_1 to $m_1 + 1$, with the exception of the crossing of one of the points where majority choice is optimal), ranging from a probability of the order 10^{-129} for $m_1 = 1$ to the order $1/100$ for $m_1 = 49$. When $m_1 > 50$, at points where minority choice is optimal, the probability decreases smoothly, with a downward jump at each

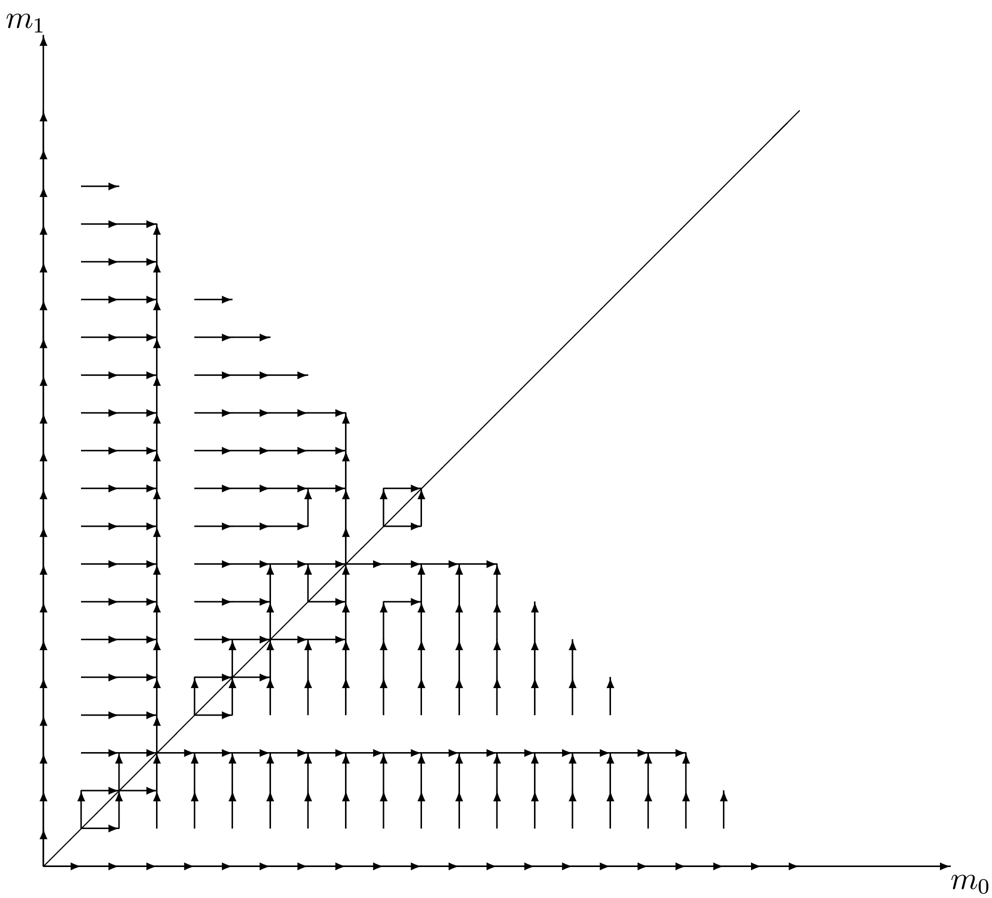

Figure 5: Optimal strategy for uninformed agents ($p = 1/4$)

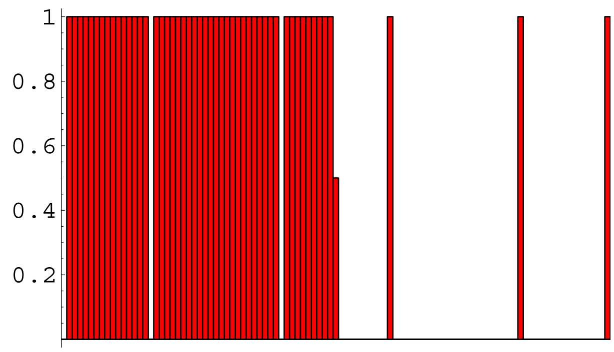

Figure 6: Optimal action of an uninformed agent at $(100 - m_1, m_1)$ as a function of m_1, $p = 1/20$

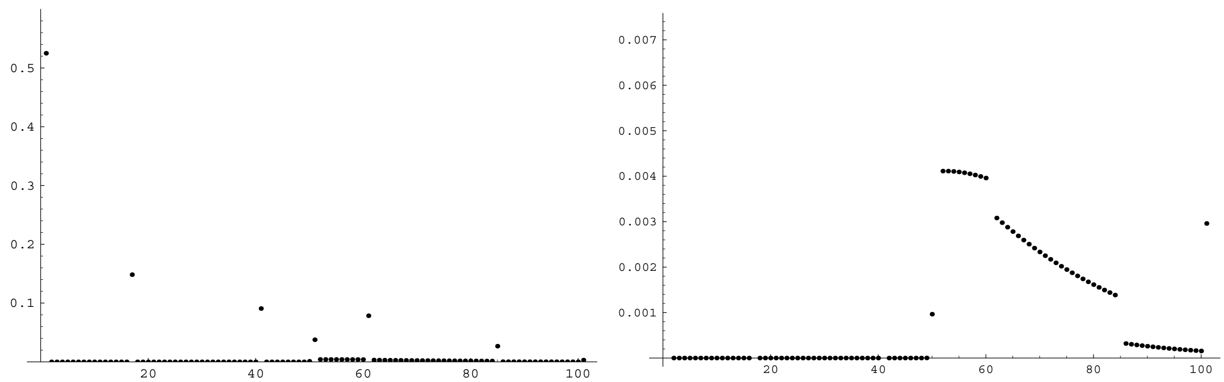

Figure 7: Probability under state ω_0 of history $(100 - m_1, m_1)$ as a function of m_1, $p = 1/20$

exceptional point, as can be observed from the right panel.

4 $p > \frac{1}{2}$: Wise Majorities?

We have so far established that minority wisdom always occurs after some history when $p < 1/2$ (in particular, after the histories $(1, 2)$ and $(2, 1)$), and indeed, after infinitely many histories provided $p < p_1$. In this section we consider the alternative case in which $p > \frac{1}{2}$, so that agents begin by following the majority should histories $(1, 2)$ or $(2, 1)$ be reached. This presages a dramatic change in the optimal strategy. Surprisingly, despite an expectation that a majority of agents are informed (and, therefore, choosing correctly), it is not always the case that following a majority is optimal. To gain some insight into this case, let us return to the simple example.

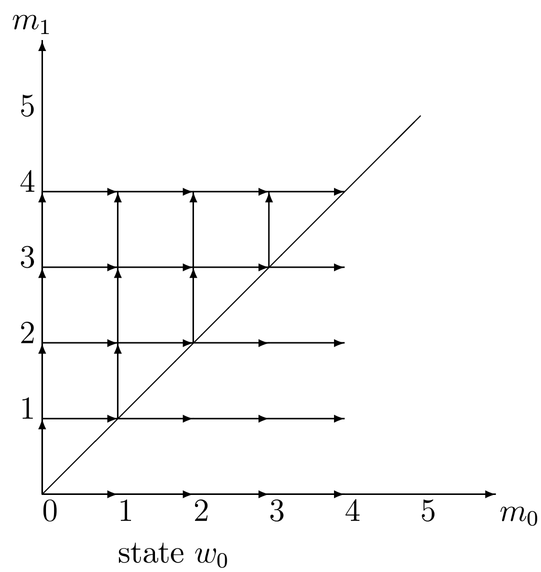

Figure 8: Paths Under Majority Choice

4.1 An Illustration: Following the Majority at $(1, 2)$

Since $p > 1/2$, we already know that majority choice is optimal after history $(1, 2)$ (i.e., $\alpha(1, 2) = 1$). Figure 8 depicts possible paths for state ω_0 if majorities are always followed (the paths for state ω_1 are a simple reflection of these). It will be useful to compute $\Pr\{2, 2 \mid \omega_0\}$ and $\Pr\{1, 3 \mid \omega_0\}$. Under ω_0, history $(2, 2)$ cannot be reached from history $(2, 1)$, since majority choice at $(2, 1)$ implies that history $(3, 1)$ necessarily follows, whether the fourth agent is informed or not. Therefore,

$$\Pr\{2, 2 \mid \omega_0\} = \Pr\{1, 2 \mid \omega_0\} \cdot p = \frac{3}{4} p^2 (1 - p)^2 .$$

To understand this, observe that, at history $(1, 2)$, only an informed agent would choose action 0. The formula for $\Pr\{1, 2 \mid \omega_0\}$ has been derived in Section 3. Consider now history $(1, 3)$ under state ω_0. Either the previous history was $(0, 3)$, in which case the last agent was informed (and all prior agents were not), or it was $(1, 2)$, in which case the last agent must have been uninformed and rationally have chosen the majority action. Therefore:

$$\Pr\{1, 3 \mid \omega_0\} = \Pr\{0, 3 \mid \omega_0\} \cdot p + \Pr\{1, 2 \mid \omega_0\} \cdot (1 - p) = \frac{5}{8} p (1 - p)^3 .$$

Observe that $\Pr\{2, 2 \mid \omega_0\} < \Pr\{1, 3 \mid \omega_0\}$ if and only if $p < 5/8$. That is, for such p, a tie is less likely than a mistaken majority, conditional on state ω_0. This may appear surprising. To understand this, recall that majority choice is optimal at $(1, 2)$ and $(2, 1)$. Under majority choice, if the majority happens to be right, it can only grow from that point on. Conversely, if a history produces a tie, truth must have been with the minority in the previous period, and the last agent must have been informed. If an uninformed agent had come along instead, the mistaken majority would have grown even larger. Thus, if p is close to $1/2$, it is hardly more likely to reach a tie from that point on, than to reach a history along which the mistaken majority exceeds the minority by two (these two events are equally likely if $p = 1/2$). But the latter history can be reached by other paths as well: for instance, it can be reached by a path in which the last agent was informed. Therefore, if p is close enough to $1/2$, the history $(1, 3)$ is more likely than history $(1, 2)$ (under ω_0). Regardless of this relationship, however, trivial calculations show that majority choice is also optimal at histories $(1, 3)$ and $(3, 1)$, provided that $p \geq 1/2$.

Why then does all this matter? Consider now history $(2, 3)$ under states ω_0 and ω_1 respectively. First,

$$\Pr\{2, 3 \mid \omega_0\} = \Pr\{1, 3 \mid \omega_0\} \cdot p + \Pr\{2, 2 \mid \omega_0\} \frac{1-p}{2},$$

since either the previous history was $(1, 3)$ and the last agent was informed (for an uninformed agent would have followed the majority), or the previous history was $(2, 2)$, and the last player was an uninformed agent who happened to make the wrong choice. Second,

$$\Pr\{2, 3 \mid \omega_1\} = \Pr\{2, 2 \mid \omega_1\} \left(p + \frac{1-p}{2}\right),$$

as the history $(2, 3)$ cannot be reached from history $(1, 3)$ if the state is ω_1, but it can be reached from history $(2, 2)$ if either the last agent was informed or uninformed but lucky.

Since $\Pr\{2, 2 \mid \omega_0\} = \Pr\{2, 2 \mid \omega_1\}$, it follows immediately that:

$$\Pr\{2, 3 \mid \omega_1\} > \Pr\{2, 3 \mid \omega_0\} \Leftrightarrow \Pr\{2, 2 \mid \omega_0\} > \Pr\{1, 3 \mid \omega_0\},$$

that is, majority choice is optimal after history $(2, 3)$ if and only if $p > 5/8$. If $p \in (1/2, 5/8)$, it is now majority choice that breaks down!

4.2 Majority Choice: Properties of the Optimal Strategy for $p > \frac{1}{2}$

The reader presumably sees a parallel with the case $p < 1/2$: for $p < 1/2$, minority choice is optimal for the first agents, but not, once large populations are considered, for some sufficiently large minorities (the lower p, the larger the necessary minority). For $p > 1/2$, majority choice is optimal for the first agents, but not necessarily, once large populations are considered, for some (as it turns out) sufficiently large minorities (the larger p, the larger the necessary minority). However, as we will see, this analogy has its limitations, as this is only valid as long as p is below some upper bound. For p sufficiently large, majority choice is optimal, no matter how large the population and no matter how small the majority. To be more formal, define the majority choice strategy as the (symmetric) strategy that necessarily mimics the choice of the majority: $\alpha^{**}(m_0, m_1) = 0$ if and only if $m_0 > m_1$.

Under majority choice, the history follows a random walk that is rather simple (as depicted in Figure 8). If, for example, the true state is ω_0, a strict majority for action 0 acts as an absorbing barrier: if action 0 achieves a majority at any point then all subsequent agents, both informed and uninformed, follow and choose action 0. In contrast, if a strict majority exists for action 1 then subsequent agents choose actions dependent on their information: informed agents follow the minority and choose action 0, and uninformed agents choose action 1. Matters are hardly more complicated at a tie. Therefore, as most agents are informed, a strict majority for action 0 obtains eventually. Using these properties of choice sequences in combination with the literature on random walks, the following rather striking conclusion emerges.

Proposition 2 *The majority choice strategy is optimal if and only if* $p \geq 7/9$.

Proof: Under the majority choice strategy, it is possible to describe quite explicitly the conditional probabilities of the various histories. Namely,

$$\Pr\{m_0, m_1 \mid \omega_0\} = \sum_{r=1}^{m_0+1} 2^{-r} \frac{m_1 - m_0 + r - 1}{m_1 + m_0 - r + 1} \binom{m_1 + m_0 - r + 1}{m_0 - r + 1} p^{m_0} (1-p)^{m_1}, \; m_1 > m_0,$$
$$\Pr\{m, m \mid \omega_0\} = \Pr\{m - 1, m \mid \omega_0\},$$
$$\Pr\{m_0, m_1 \mid \omega_0\} = \frac{1+p}{2} \Pr\{m_1, m_1 \mid \omega_0\}, \; m_1 < m_0,$$

The first formula follows from the formula for the number $N(m_0, m_1)$ of paths going from $(0,0)$ to (m_0, m_1) touching exactly r times the horizontal axis without ever crossing it (see Mohanty (1979), formula (4.9)):

$$N(m_0, m_1) = \frac{m_1 - m_0 + r - 1}{m_1 + m_0 - r + 1} \binom{m_1 + m_0 - r + 1}{m_0 - r + 1}.$$

The other two formulas are obvious. Trite computations show that, for $n \in 2\mathbb{N}$:

$$\frac{\Pr\{m, m \mid \omega_0\}}{p^m (1-p)^m} = \frac{C(2; 2m)}{2^{2m}} := \sum_{r=0}^{2m-1} (-1)^r 2^{-r} C_{2m-1-r} + (-1)^{2m} 2^{-2m},$$

where $C_n = \binom{2n}{n} / (n+1)$ is the Catalan number ($C(2; n)$ is known as a generalized Catalan number).

Let $p(m_0, m_1) = \Pr\{m_0, m_1 \mid \omega_0\}$. For $p \geq 3/4$, the probability $p(m_0, m_1)$ is non-increasing in m_1, for $m_1 > m_0$. To see this, observe that $p(m_0, m_1 + 1)$, $p(m_0, m_1)$ are given by the aforementioned summations, and consider the ratio of the corresponding summands:

$$
\begin{aligned}
\frac{N(m_0, m_1+1)}{N(m_0, m_1)}(1-p) &\leq \frac{1}{4}\frac{N(m_0, m_1+1)}{N(m_0, m_1)} \\
&= \frac{1}{4}\frac{m_1+1-(m_0+1-r)}{m_1-(m_0+1-r)}\frac{m_1+(m_0+1-r)}{m_1+1} \\
&\leq \frac{1}{4}\frac{r+1}{r}\frac{2m_0+2-r}{m_0+2} \leq \frac{1}{4}4 = 1,
\end{aligned}
$$

where the second inequality follows by observing that the expression is decreasing in m_1, and the last one from the fact that the expression is decreasing in $r \leq m_0+1$.

It follows that majority choice is optimal provided that:

$$
\Pr\{(m+1, m) \mid \omega_0\} > \Pr\{(m, m+1) \mid \omega_0\}.
$$

We will show that this inequality holds for all m if and only if $p \geq 7/9$. The previous inequality is equivalent to:

$$
\frac{C(2; n+1) - C(2; n)}{2C(2; n)} \leq \frac{p}{1-p},
$$

for $n := 2m$. As can be readily verified, $C(2; n+1)/C(2; n)$ is increasing in n and bounded, and converges therefore to some limit l. It is clear, from the definition of $C(2; n)$, that:

$$
\begin{aligned}
C(2; n+1) + C(2; n) &= 2^{n+1}C_n, \text{ and thus also} \\
C(2; n+2) - C(2; n) &= 2^{n+2}C_{n+1} - 2^{n+1}C_n.
\end{aligned}
$$

>From the identity:

$$
\frac{C(2; n+2) - C(2; n+1)}{C(2; n+1)} = \frac{C(2; n+2) - C(2; n)}{C(2; n+1)} - \frac{C(2; n+1) - C(2; n)}{C(2; n)}\frac{C(2; n)}{C(2; n+1)},
$$

it follows that:

$$
\begin{aligned}
l &= \lim_n \frac{C(2; n+2) - C(2; n)}{C(2; n+1) + C(2; n)} = \lim_n \frac{2^{n+2}C_{n+1} - 2^{n+1}C_n}{2^{n+1}C_n} \\
&= 2\lim_n \frac{C_{n+1}}{C_n} - 1 = 2\lim_n \left(4 - \frac{6}{n+2}\right) - 1 = 7.
\end{aligned}
$$

Hence,

$$
\frac{C(2; n+1) - C(2; n)}{2C(2; n)} \leq \frac{p}{1-p} \forall n \Leftrightarrow p \geq \bar{p} := 7/9.
$$

It is clear that the optimal action given some event (m_0, m_1) only depends on the optimal actions at all events (m_0', m_1') for $m_1' < m_1$, and $m_0' \leq m_0$. It follows that there are thresholds p_i such that the majority rule is optimal as long as the minority does not exceed i, provided $p \geq p_i$. The threshold p_i is increasing in i and tends to $\bar{p}$ as $n \to \infty$. ■

The optimal strategy for $p \geq \frac{7}{9}$ is, therefore, straightforward. For $p \in \left(\frac{1}{2}, \frac{7}{9}\right)$, however, the optimal strategy (once majority choice breaks down) is complicated and irregular, as is the case for $p < \frac{1}{2}$.

5 Efficiency

Although minority choice allows undesirable herds to be avoided, it does not aggregate information efficiently over time as it runs into the fundamental problem that to be effective it must ultimately stop. If minority choice were always optimal, and informed agents were rare ($p < \frac{1}{2}$), then provided only the first agent made an incorrect choice the tally would tend to oscillate around a tie, with larger departures from ties being caused by strings of informed agents, eventually cancelled out by the larger numbers of uninformed agents that, choosing with the minority, would cause a reversion to a tie. The optimal strategy would, in effect, jam the signals of prior agents.

But we already know that minority choice must break down, sooner or later, independently of the value of p. The question of efficiency becomes therefore nontrivial. What is the limit of the expected fraction of agents who choose correctly, as the number of agents grows large? Given the optimal strategy, define X_n as the random variable that corresponds to the choice of the nth agent: $X_n = 0$ if he chooses 0, $X_n = 1$ otherwise. Define also:

$$S_n := \sum_{i=1}^{n} X_n \text{ and } M_n := S_n/n.$$

The optimal strategy is *asymptotically efficient* if $\lim_{n\to\infty} \mathbb{E}\left[M_n \mid \omega_0\right] = 0$, where $\mathbb{E}\left[\cdot\right]$ denotes expectations. [Since it is symmetric, this also implies that $\lim_{n\to\infty} \mathbb{E}\left[M_n \mid \omega_1\right] = 1$.] A strategy is *efficient* if it minimizes $\mathbb{E}\left[M_n \mid \omega_0\right]$ and $\mathbb{E}\left[1 - M_n \mid \omega_1\right]$, for all n (we could have equivalently stated this definition in terms of S_n rather than M_n). At first glance, it may not be obvious that there exists a strategy that simultaneously minimizes the expected number of incorrect decisions for all population sizes. For example, why wouldn't it be in the collective interest to have a group of early agents choose against their own interest for the sake of avoiding signal-jamming? It is nevertheless the case that efficient strategies exist. One of them is described in the proof of the following result, based on an application of the welfare improvement principle (see Banerjee and Fudenberg (2004), Smith and Sørensen (1999)).

Proposition 3 *The optimal strategy is asymptotically efficient, but not efficient.*

Proof: Consider the probability that agent n makes the incorrect decision (*before* he observes his private signal) if he follows the strategy of either following his informative signal, if the case occurs, or mimicking one of his predecessors at random. [The optimal strategy is the one that minimizes this probability.] While this is not a particularly bright strategy, it nevertheless guarantees that this probability does not exceed:

$$q_n = (1-p)\frac{\sum_{j=1}^{n-1} q_j}{n-1},$$

where q_j denotes the corresponding probability for agent j, with the understanding that $q_1 = (1-p)/2$. [Of course, q_j is an upper bound on the probability agent $j < n$ takes an incorrect decision, since agent j follows the optimal strategy.] Solving, we get, for all n:

$$q_n = \frac{1-p}{2}\prod_{j=1}^{n-1}\left(1-\frac{p}{j}\right).$$

Taking logarithms, the convergence of this sequence is equivalent to the divergence of the series $\ln(1-p/n)$, which follows from the divergence of the series $-p/n$.

Thus, the probability of an incorrect decision tends to 0 along any path. One efficient strategy is: "choose action 0 until at least one agent has chosen action 1 so far" (obviously, there is another one, where the role of 0 and 1 are exchanged). Of course, informed agents follow their information. The argument is rather simple. Consider any agent n, uninformed. If one or more of his predecessors are informed, he is sure to choose the correct action (since either the correct action is 1, and therefore someone has chosen 1 before, so agent n chooses 1; or the correct action is 0, in which case nobody has chosen 1, and agent n chooses 0 as well). If none of his predecessors is informed, then indeed, both actions are equally good, and 0 is optimal. For n agents, the expected number of agents taking the incorrect action under the efficient strategy is:

$$S_n^* = \frac{1-p}{2p}\left(1-(1-p)^n\right),$$

which tends to $(1-p)/(2p)$.

So the total expected number of agents taking an incorrect action is finite under the efficient strategy. The equilibrium strategy is not efficient, because, in some circumstances, while the deciding agent knows that at least one of his predecessors was informed, he is not able to infer what this information was. For instance, if the fourth agent observes two choices for 1 and one choice for 0 – which implies that at least one agent was informed – it could be that the sequence was $(1,0,1)$, with the second agent being informed of state 0, or it could be that the sequence was $(0,1,1)$, in which case the second agent was informed of state 1.

This implies that, among the first three agents, the expected number of agents choosing incorrectly for, say, $p = 1/5$, is $126/125 > 1$, while under the efficient strategy, only $122/125 < 1$ choose incorrectly. (Suboptimality also obtains for larger p: for $p = 8/9$, the numbers are $95/1458$ and $91/1458$, respectively.) ■

While the optimal strategy is not efficient – in fact, for small p, the expected number of agents taking an incorrect action is larger than under an efficient strategy by a factor that is (arbitrarily?) large – it appears to be the case, from all numerical calculations that we made, that the expected number of incorrect decisions remains bounded, for any $p > 0$. [Clearly, there exists sample paths for which the number of incorrect decisions is arbitrarily large.] This conjecture is easily proven for the case $p > \frac{1}{2}$, but we have not been able to establish it otherwise.[4] As mentioned in the proof of the previous result, this limit is finite for any efficient procedure. However, the proportional sampling scheme used there to prove that $\mathbb{E}[M_n \mid \omega_0]$ tends to zero under the equilibrium strategy does not yield that $\mathbb{E}[S_n \mid \omega_0]$ converges. To the contrary, under this sampling scheme, this number diverges, as is readily verified. The next figure illustrates this discussion (where S_n^* represents the efficient strategy and S_n the optimal strategy).

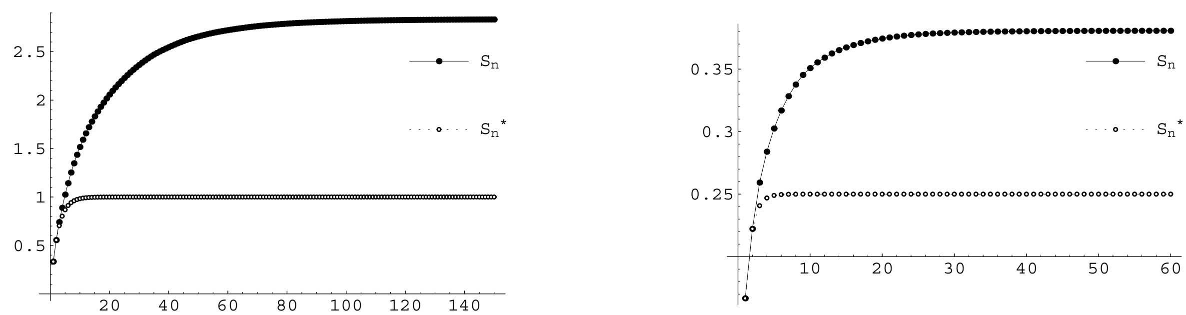

Figure 9: Expected number of incorrect decisions, (i) $p = 1/3$ and (ii) $p = 2/3$

6 Discussion

6.1 The tie-breaking rule

The structure of the optimal strategy is sensitive to the tie-breaking rule that is adopted. The qualitative features of the optimal strategy obtain for any tie-breaking rule, as long as agents that are indifferent do not systematically choose a pure action. For instance, if indifferent agents systematically choose action 0, then the optimal strategy is the efficient strategy described above: as soon as an agent chooses action 1, all his followers do the same (on the equilibrium path), otherwise uninformed agents choose action 0. Since this strategy is efficient, why shouldn't agents use this tie-breaking rule rather than the symmetric one we have used?

[4] If $p > 1/2$, observe that, by following the majority decision, the nth agent can secure a probability of error not exceeding $(1-p)\,q_{\lceil n/2 \rceil}$, where q_j is the jth agent's probability of error under that method, and $\lceil x \rceil$ is the smallest integer no smaller than x. This implies that the sum of probabilities of errors does not exceed $\sum_{j=1}^{\infty} \frac{1-p}{2} 2^j (1-p)^j = (1-p)^2 / (2p-1)$, which is indeed finite.

What matters for our results is not that agents randomize, but that they are unsure how uninformed agents decide when they face a tie. There are two obvious reasons for why such uncertainty is plausible. First, we have assumed that preferences are common knowledge. Second, we have assumed that less informed agents are completely uninformed. If either assumption fails, as is the case if, for instance, an agent's utility has an idiosyncratic component, no matter how small, that is private information, or if less informed agents receive noisy, but neither completely uninformative nor perfectly correlated signals, the necessary uncertainty would obtain.

Therefore, we view our model as a simplification of a richer model in which agents have different preferences and are neither perfectly informed nor uninformed. The results of such a richer model are described in the following subsections. Simple calculations, as in Section 3, show that minority choice arises in these broader environments. Rather than repeat the tedious detail, we take the opportunity at several points to instead focus on optimal behavior further down the decision sequence, showing patterns of behavior for the 101st agent.

6.2 Differentially informed agents

Suppose now that some agents simply receive a better signal than others and that all agents are potentially informed. That is, conditional on the state, a more informed agent receives the correct signal with probability q, while a less informed agents receive the correct signal with probability r, where $1/2 \leq r \leq q \leq 1$. How does the optimal strategy vary with our three parameters: p, q, and r? While we are unable to provide a symbolic characterization for arbitrary histories, the actual computations can be carried out for any set of specific values.

The extreme case of homogeneous agents is standard. If $q = r$, the optimal strategy of an agent who receives a signal in favor of 1 is as follows: $\alpha(m_0, m_1) = 0$ if $m_0 > m_1 + 1$, $= \frac{1}{2}$ if $m_0 = m_1 + 1$, and $= 1$ otherwise. In particular, it is independent of the quality of the signal. So majority choice is optimal in this case, at least when the majority exceeds the minority by more than one. The reasoning is familiar. The first agent chooses according to his signal. The second, knowing this, is indifferent if his signal is opposed to the first agent's choice (he then randomizes), and chooses according to his signal otherwise. Informational cascades can take place from agent three onwards: if the first two agents picked the same action, the third agent ignores his signal and follows the majority. Otherwise, he follows his signal. Etc. Heterogeneity is therefore a necessary ingredient of our results.

Turn now to heterogeneous, although not necessarily extreme, signals and consider the optimal strategy of the agent who comes in position 101, with a signal of accuracy r that is in favor of state ω_1. Suppose, as before, that informed agents are perfectly informed ($q = 1$). How does his action vary, as a function of m_1, the number of agents who have chosen action 1 so far? This depends on the value of r, the quality of his signal; Figure 10 illustrates this effect for several values of r (where, as before, the value represented is the probability of choosing action 1; note that for $r \neq \frac{1}{2}$ the figures need not be symmetric as less informed agents now possess private information). The case of $r = 1/2$, in which the agent's signal is uninformative, is that considered previously and following the minority choice is generally optimal, with only two exceptions (see panel 1 of the figure). As can be seen from the remaining panels, minority choice is less frequent for larger values of r, but it does persist. These findings establish two conclusions: that uninformed agents are not necessary for minority wisdom to arise, and that the relative wisdom of a minority can be significant (and sufficient to overcome a private signal of $\frac{4}{5}$ accuracy).

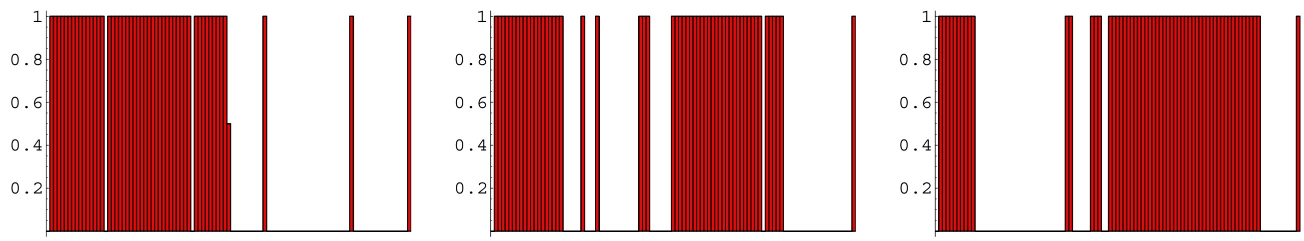

Figure 10: Optimal strategy of a less informed agent with a signal in favor of 1, as a function of m_1, $m_0 + m_1 = 100, p = 1/20, q = 1$, (i) $r = 1/2$, (ii) $r = 3/5$ and (iii) $r = 4/5$

Figure 11 shows how the 101st agent's decision varies with the fraction of better informed agents, still assuming that better informed agents are perfectly informed, and setting r, the accuracy of less informed agents, to 3/5. The effect of a higher p seems to be to increase the "volatility" of his decision: when very few agents are likely to be perfectly informed, he follows the minority and goes against his own signal, unless this minority is large enough. As the expected fraction of informed agents increases, the agent appears more likely to choose with the majority, although a significant amount of minority choice persists and his decision becomes very sensitive to the specific size of the majority.

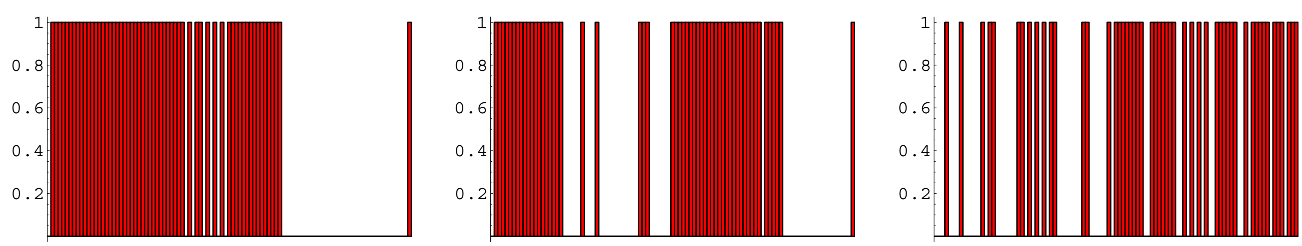

Figure 11: Optimal strategy of a less informed agent with a signal in favor of 1, as a function of m_1, $m_0 + m_1 = 100, q = 1, r = 3/5$, (i) $p = 1/50$, (ii) $p = 1/20$ and (iii) $p = 1/2$

Finally, Figure 12 illustrates the effect of increasing the quality of the information received by the more informed agents. Setting $q = r$ is again the traditional model and majority choice prevails (except when outweighed by private information). The effect for $q > r$ seems to be what one would expect given previous results: the higher the quality of information available to better informed agents, the more likely it is that minority choice is optimal. These figures also establish that unbounded private signals are not necessary for minority wisdom to arise.

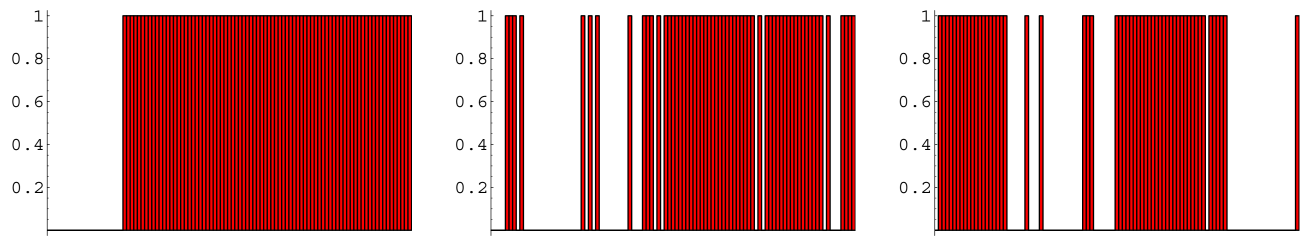

Figure 12: Optimal strategy of a less informed agent with a signal in favor of 1, as a function of m_1, $m_0 + m_1 = 100, p = 1/20, r = 3/5$, (i) $q = 7/10$, (ii) $q = 9/10$ and (iii) $q = 1$

The quality of information available to agents can be indexed by $\sigma = pq + (1-p)\,r$. Fixing n and two out of the three parameters p, q and r, one can verify numerically that increasing the quality of information, as measured by the third parameter, tends to decrease the expected number of mistakes.[5] Holding σ fixed, it is also the case that increasing the spread between the quality of the signals received by well-informed and less informed agents is socially beneficial, as it tends to decrease the expected number of mistakes as well.

6.3 Continuous Beliefs

The calculations of the previous examples depend on the restriction to two signals, although the logic of the results do not. The possibility for wise minorities arises even with a continuum of signals and beliefs with full support. For example, suppose that each agent gets an independent draw t from some distribution f with support $\left[\frac{1}{2}, 1\right]$; the agent is informed of the realized value of t and then given a signal of the state of the world that is correct with probability $t \geq \frac{1}{2}$, and incorrect with probability $1 - t \leq \frac{1}{2}$. Let the density of f be given by

$$f\,(t; \alpha, p) = 2\,(1+\alpha)\left((1-p)\,(2\,(1-t))^{\alpha} + p\,(2t-1)^{\alpha}\right),$$

where $\alpha \geq 1$ and $0 \leq p \leq 1$. This density covers a broad array of possibilities: for $\alpha = 1$ and $p = \frac{1}{2}$ it corresponds to the uniform distribution, and when $\alpha \to \infty$, f converges to the baseline model of the current paper with two point support $\left\{\frac{1}{2}, 1\right\}$ and respective mass $1-p$ and p. Roughly speaking, α describes the convexity of the density (convexity is increasing in α), and q in this context represents "tilt," where lower p puts more weight on uninformative signals and values of p near one puts more weight on informative signals (and so p's role here is analogous to that in the baseline two signal model).

[5] It is possible to find pairs of values (sufficiently close to each other) that violate the monotonicity, but these counterexamples are essentially driven by integer problems.

Smith and Sørensen (1997) study the uniform case and find that minority wisdom does not arise, thereby proving that heterogeneity of signals is not sufficient for minority wisdom. Examining agent beliefs at history $(2,1)$, it can be shown that the minority is wiser than the majority for sufficiently small p and large α; that is, as in the binary case, if agents are more likely to receive relatively inaccurate signals and sufficient mass is toward the extremes of the distribution (both informed and uninformed).[6] Thus, sufficient variance of beliefs is necessary for the decisions of a minority to convey more information than those of the majority. The left panel of Figure 13 depicts for the 4th agent the likelihood ratio of the two states given the history $(2,1)$ as p and α vary. Of most interest is the size of this ratio relative to one as, modulo his private signal, the agent wishes to follow the minority if the ratio is greater than one, and the majority otherwise; for ease of visualization, this sign is depicted in the right panel of the figure, where a value of one corresponds to a likelihood ratio greater than one and a value of negative one corresponds to a value less than one.

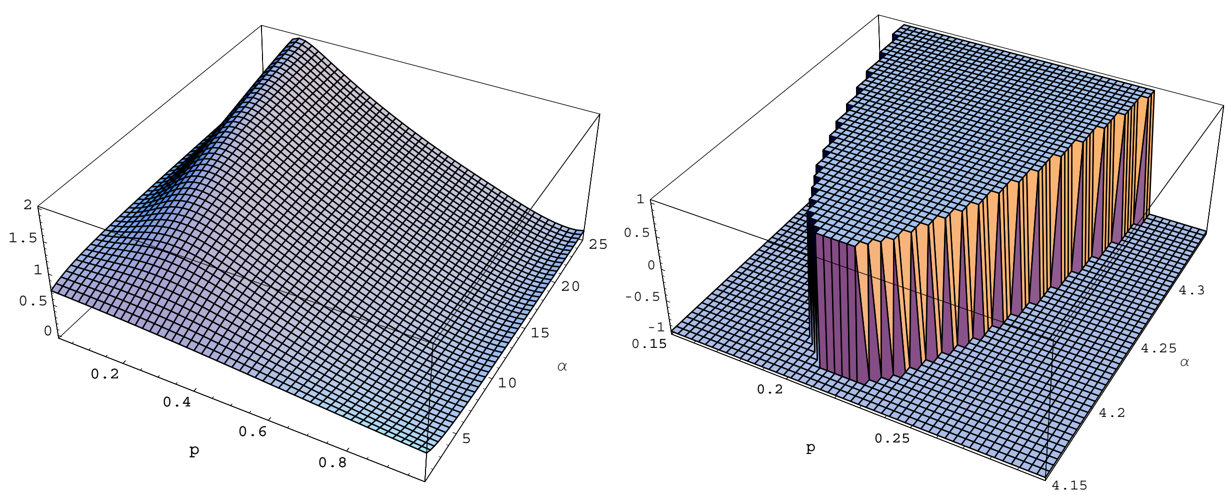

Figure 13: (i) Likelihood ratio $\frac{\Pr\{(2,1)|\omega_1\}}{\Pr\{(2,1)|\omega_0\}}$, and (ii) the sign of its value relative to one.

Notable in the right panel is the non-monotonicity in p of agent 4's optimal response for some values of α. In the baseline binary signal model, the probability of breaking unanimity at any point, conditional on these choices being for the wrong state, is always the same and equal to p, the probability that an agent is informed. With continuous beliefs this property no longer holds and it is decreasingly likely that unanimity is broken as the number of previous choices

[6] One may be interested in the condition that delineates the domain of minority wisdom. An intriguing conjecture by Peter Norman Sørensen is that log-concavity of the log likelihood ratio (of the distribution of unconditional beliefs) is sufficient to rule out wise minorities. Numerical verification of the distribution considered here establishes that log-concavity fails for all values of α larger than one. Thus, Sørensen's conjecture is not disproved by our calculations and remains open.

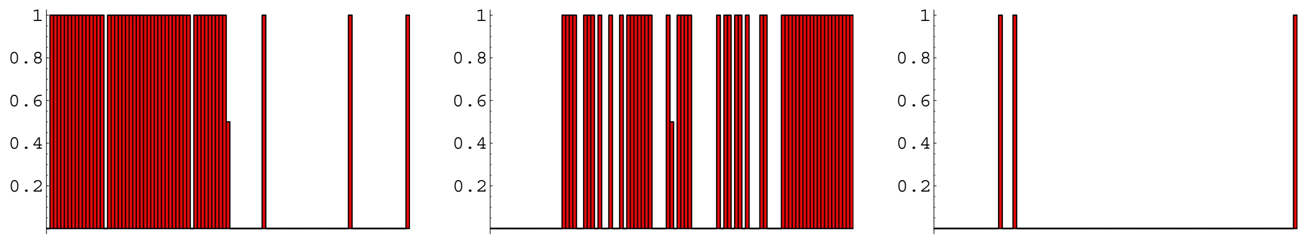

Figure 14: Optimal strategy of an uninformed agent, as a function of m_1, $m_0 + m_1 = 100, p = 1/20$, (i) no partisans, (ii) 10% partisans for each decision, (iii) 10% partisans for action 1, none for action 0

increases. Consequently, although lower p means that the path to $(2, 1)$ through $(1, 1)$ conveys less information in favor of the majority, it also implies that this path is relatively more likely than that through $(2, 0)$ (which conveys information in favor of the minority). These dual forces interact in such a way that the wisdom of the minority may not be monotonic in p.

6.4 Partisans

Let us return to the case in which signals are either perfectly informative, or not informative at all, and suppose that some agents are *partisans*: that is, they have a preference for one of the actions, so that it is a dominant strategy for them to choose that action independently of the history that precedes. Whether a player is a partisan or not is private information, and each agent has a positive, commonly known, probability of being partisan, which is independent of the state of nature and of the history. We shall now refer to an uninformed agent as one that is neither partisan nor informed. Partisanship introduces yet another source of uncertainty for uninformed agents, who must account for the possibility that some decisions may not have been taken on the basis of direct or indirect evidence, but simply because of the agent's preferences. We illustrate in Figure 14, for some specific parameters, how the existence of such partisans affects an uninformed agent's decision, for the cases in which partisans are equally likely to choose one or the other decision, and in which there are only partisans for decision 1.

Introducing partisans has a strong impact on the decision of uninformed agents. A minority need not be the expression of the few informed agents any longer. Rather, it may bring together those agents who favor the action against which all evidence pleads. Indeed, as can be observed from the case in which both types of partisans exist, while minority choice still persists, it can now be found only for sufficiently large minorities. In aggregate, however, the optimal strategy appears to be at least as complex as in the case without any partisans.

If partisans are all expected to choose action 1 (case (iii)), then the effect on behavior is even more dramatic: in almost every contingency, the uninformed agent prefers to choose action 0. In fact, this behavior balances out the presence of partisans. Under state ω_0, the expected number of agents who choose action 1, among the first hundred agents, is 20.97 in the absence of partisans (for the parameter values used in Figure 14). With partisans on either side (case (ii)), this number almost doubles to 40.77. That is, partisans mist the inference problem, and out of the eighty or so non-partisans, more than thirty take the incorrect decision. If all partisans favor action 1 (case (iii)), this expected number goes down to 21.45, which means that, among the ninety or so non-partisans, hardly more than ten got it wrong![7] The benefit is not without a corresponding cost, however, which arises in state ω_1 when the expected number of agents to choose action 1 is only 64.59 and more than 35 agents take the incorrect action. It would seem, therefore, that partisans are useful for information aggregation only when they are wrong.

The inclusion of one-sided partisans introduces an effect not dissimilar to that within the efficient strategy of Section 4, although asymmetrically and less effectively. The assumption that 10% of agents are partisans for action 1 makes it more likely that the first agent chooses action 1 rather than action 0, thereby making later observations of action 0 more informative. That is, as agents expect early agents to choose 1, any choice of action 0 most likely reflects an informed decision worthy of imitation (as there are no action 0 partisans). On the other hand, and unlike in the efficient strategy, uninformed agents may choose action 0 even in state ω_1. By the preceding logic, this induces later uninformed agents to follow with action 0, perpetuating an incorrect herd that otherwise may have been overcome in the absence of action 1 partisans.

7 Conclusion

Although becoming a prominent topic of economic research only in the past decade and a half, social learning has long concerned thinkers from a variety of fields. Indeed, the question of minority wisdom that is at the heart of the current paper was addressed insightfully by Kierkegaard in the 19th century:

> "Truth always rests with the minority, and the minority is always stronger than the majority, because the minority is generally formed by those who really have an opinion, while the strength of a majority is illusory, formed by the gangs who have no opinion and who, therefore, in the next instant (when it is evident that the minority is the stronger) assume its opinion ... while Truth again reverts to a new minority."
> – Søren Kierkegaard (1813-1855).

Although it is likely that Kierkegaard did not have a specific model in mind (if any at all), we show here that he grasped a degree of truth. Key to our results is the assumption that agents are unaware of the sequence in which previous decisions were made. To infer the private information of others, therefore, agents must first draw inferences about what previous agents knew when making their decisions. This decision problem, when combined with differentially

[7] In case (i), among the first *ninety* agents, the expected number making the incorrect decision is 19.71.

informed agents, not only leads to different behavior than would otherwise arise, but induces the unexpected situation in which the minority is more likely than the majority to be right.

The model we analyze is far from capturing real decision environments although the main findings appear robust to simple generalizations. The possibility that sequence matters to individual behavior suggests that more study is needed of how history affects behavior and the informativeness of aggregate statistics. The opportunities are broad for application of these ideas, both in the laboratory and in the field.

References

[1] Banerjee, A. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107, 3 (August): 797-817.

[2] Banerjee, A. 1993. "The Economics of Rumours." *Review of Economic Studies* 60, 2 (April): 309-327.

[3] Banerjee, A. and D. Fudenberg. 2004. "Word-of-mouth learning." *Games and Economic Behavior* 46: 1–22.

[4] Bikhchandani, S., Hirshleifer, D. and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades." *Journal of Political Economy* 100, 5 (October): 992-1026.

[5] Çelen, B. and S. Kariv. 2004. "Observational Learning Under Imperfect Information." *Games and Economic Behavior* March, 47(1): 72-86.

[6] Chamley, C. 2004. *Rational Herds: Economic Models of Social Learning*. Cambridge University Press: Cambridge, UK.

[7] Kierkegaard, S. *The Diary of Søren Kierkegaard*, pt. 5, sect. 3, no. 127, entry from 1850, ed. by Peter Rohde, 1960.

[8] Mohanty, G. 1979. *Lattice Path Counting and Applications*. Academic Press: London, U.K.

[9] Smith, L. and P. Sørensen. 1997. "Martingales, Momentum, or Mean Reversion? On Bayesian Conventional Wisdom." Working paper.

[10] Smith, L. and P. Sørensen. 1999. "Rational Social Learning by Random Sampling." Working paper, University of Michigan.

[11] Smith, L. and P. Sørensen. 2000. "Pathological Outcomes of Observational Learning." *Econometrica* 68 (March): 371-398.

[12] Smith, L. and P. Sørensen. 2001. "Informational Herding and Optimal Experimentation." working paper.

Appendix:

Proof of Proposition 1, part (i): The following remarks will prove useful in the argument.

Remark 1: The problem is obviously symmetric: if one action is optimal when m_1 out of n agents have taken action 1 so far, then the other action is optimal when $n - m_1$ out of n agents have taken that action so far. Formally:

$$\alpha(m_0, m_1) = 1 - \alpha(m_1, m_0).$$

Remark 2: Let $\Pr\{\omega_i \mid m_0, m_1\}$ denote the probability that the state is $\omega_i \in \Omega$, given that m_i agents have taken action $i = 0, 1$, and let $\Pr\{m_0, m_1 \mid \omega_i\}$ denote the probability that m_i agents take action i, given state $\omega_i \in \Omega$ (conditional, of course, on $m_0 + m_1$ agents exactly having chosen). It follows from Bayes rule that:

$$\begin{aligned}
\Pr\{\omega_i \mid m_0, m_1\} &= \frac{\Pr\{m_0, m_1 \mid \omega_i\} \Pr\{\omega_i\}}{\Pr\{m_0, m_1 \mid \omega_0\} \Pr\{\omega_0\} + \Pr\{m_0, m_1 \mid \omega_1\} \Pr\{\omega_1\}} \\
&= \frac{\Pr\{m_0, m_1 \mid \omega_i\}}{\Pr\{m_0, m_1 \mid \omega_0\} + \Pr\{m_0, m_1 \mid \omega_1\}}.
\end{aligned}$$

Therefore,

$$\Pr\{\omega_0 \mid m_0, m_1\} > \Pr\{\omega_1 \mid m_0, m_1\} \Leftrightarrow \Pr\{m_0, m_1 \mid \omega_0\} > \Pr\{m_0, m_1 \mid \omega_1\}.$$

The advantage of working with the probabilities $\Pr\{m_0, m_1 \mid \omega_i\}$ is that they obey simple recursions, namely:

$$\begin{aligned}
\Pr\{m_0, m_1 \mid \omega_0\} &= \Pr\{m_0, m_1 - 1 \mid \omega_0\} \cdot (1 - p)\, \alpha(m_0, m_1 - 1) \\
&\quad + \Pr\{m_0 - 1, m_1 \mid \omega_0\} \cdot (p + (1 - p)(1 - \alpha(m_0 - 1, m_1))), \\
\Pr\{m_0, m_1 \mid \omega_1\} &= \Pr\{m_0, m_1 - 1 \mid \omega_1\} \cdot (p + (1 - p)\, \alpha(m_0, m_1 - 1)) \\
&\quad + \Pr\{m_0 - 1, m_1 \mid \omega_1\} \cdot (1 - p)(1 - \alpha(m_0 - 1, m_1)),
\end{aligned}$$

with boundary condition $\Pr\{0, 0 \mid \omega_i\} = 1\ \forall \omega_i \in \Omega$.

The general argument is by induction, assuming that it holds for $m_1 \leq M - 1$, and proving it holds then for $m_1 = M$ as well, for $p \leq 4^{-M}$. For $m_1 = 1$, it was established in Section 3 that the conclusion obtains provided $p \leq \left(5 - \sqrt{17}\right)/2$. Conditional on state ω_0, if 1 out of n agents have taken action 1, no more than 2 agents among the n agents could have been uninformed.

Suppose now that the proposition holds for all $m \leq M - 1$, as well as the following claim, valid for $m = 1$: conditional on state ω_0, if $m \leq M - 1$ out of n agents have taken action 1 (where m and n satisfy the assumptions of Proposition 1), then no more than $2m$ agents, out of these n agents, could have been uninformed. We establish these two claims for $m = M$ by induction on the number of agents, n. Suppose that $n = 2M + 1$ (the smallest number of agents that must be considered given the assumptions of Proposition 1). For the second claim, we must show that no more than $2M$ were uninformed, that is, at least one agent was informed. This, however, is obvious, since unanimity would obtain if all agents were uninformed. For the first claim, observe that:

$$\Pr\{M+1, M \mid \omega_1\} = \frac{1-p}{2}\Pr\{M, M \mid \omega_1\} + \Pr\{M+1, M-1 \mid \omega_1\}.$$

To see this, observe that either M out of the first $2M$ agents had taken action 1, in which case, conditional on state 1, only an uninformed agent could take action 0 (furthermore, only with probability $\frac{1}{2}$), or $M-1$ out of the first $2M$ agents had done so, in which case, even an uninformed would have taken action 1 (by the induction hypothesis on M). Similarly:

$$\Pr\{M+1, M \mid \omega_0\} = \left(\frac{1-p}{2} + p\right)\Pr\{M, M \mid \omega_0\} + (1-p)\Pr\{M+1, M-1 \mid \omega_0\}.$$

Therefore:

$$\Pr\{M+1, M \mid \omega_1\} > \Pr\{M+1, M \mid \omega_0\}$$
$$\Longleftrightarrow$$
$$\Pr\{M+1, M-1 \mid \omega_1\} > p\Pr\{M, M \mid \omega_0\} + (1-p)\Pr\{M+1, M-1 \mid \omega_0\},$$

since $\Pr\{M, M \mid \omega_i\}$ is independent of i. Observe that, if M out of $2M$ agents have taken action 1, at least one of them must have been informed. And if $M-1$ out of $2M$ agents have taken action 1, then, as no more than $2(M-1)$ could have been uninformed, at least two of them must have been informed. Therefore:

$$\begin{aligned} & p\Pr\{M, M \mid \omega_0\} + (1-p)\Pr\{M+1, M-1 \mid \omega_0\} \\ \leq\; & p\sum_{i\geq 1}\binom{2M-1}{i}\frac{1}{2}p^i(1-p)^{2M-i} + (1-p)\sum_{i\geq 2}\binom{2M-1}{i}\frac{1}{2}p^i(1-p)^{2M-i} \\ <\; & \frac{1}{2}p^2(1-p)^{2M-1}4^M, \end{aligned}$$

as long as $p < 1/2$. [The fact that the first agent to act must have been uninformed has been used in the binomial coefficient.] In addition:

$$\Pr\{M+1, M-1 \mid \omega_1\} = \frac{1}{2}p(1-p)^{M-1}.$$

Therefore, the desired inequality holds for $p \leq 4^{-M}$.

Suppose now that both claims hold for some $n \geq 2M+1$, and suppose that M out of $n+1$ agents have taken action 1. As for the second claim, either $M-1$ out of the first n agents had taken action 1, in which case no more than $2(M-1)+1 < 2M$ out of the $n+1$ agents can be uninformed, or M out of the first n agents had taken action 1. In that case, however, the $(n+1)^{th}$ agent must have been informed, as, by the induction hypothesis, an uninformed agent would have taken action 1. Therefore, in this case as well, no more than $2M$ out of the first $n+1$ agents can be uninformed. In state ω_0 at least M agents are informed; therefore:

$$\Pr\{n+1-M, M \mid \omega_0\} < \sum_{i \leq 2M} \binom{n}{i} \frac{1}{2} (1-p)^i p^{n+1-i}.$$

Observe also that:

$$\Pr\{n+1-M, M \mid \omega_1\} = \frac{1}{2} p (1-p)^{n+1-M}.$$

A sufficient condition to establish $\Pr\{n+1-M, M \mid \omega_1\} > \Pr\{n+1-M, M \mid \omega_0\}$ is, therefore,

$$\frac{1}{2} p (1-p)^{n+1-M} \geq \frac{(1-p)^{2M}}{2} p^{n+1-2M} 2^n,$$

which reduces to:

$$1 \geq p^{n-2M} (1-p)^{3M-(n+1)} 2^n.$$

The right-hand side is decreasing in n for $p \leq 4^{-M}$; and for $n = 2M+2$ (the smallest value), the right-hand side is bounded by:

$$p^2 (1-p)^{M-3} 2^{2M+2} \leq 4^{-M+1} (1-p)^{M-3} < 1.$$

The result follows. ■