Information Sampling, Belief Synchronization
and Collective Illusions

Jerker Denrell & Gaël Le Mens*

Forthcoming in Management Science (August 4, 2015)

Abstract

We demonstrate that a sampling-based mechanism can offer an alternative explanation for belief synchronization in social groups and the persistence of collective illusions. Our model assumes that people are more likely to sample popular alternatives than unpopular alternatives. We show that this mechanism is sufficient to explain belief synchronization: a strong majority of opinions will likely emerge in favor of one alternative. The reason is that the group is unlikely to move away from a state in which one alternative is very unpopular. If by chance most people come to dislike alternative A, they are all unlikely to sample it again and their opinions of A remain the same. When A is in fact the best alternative, a collective illusion has emerged because people mistakenly believe that a suboptimal alternative is the best. Our model implies that such a collective illusion is persistent. The model thus offers an existence proof that a collective illusion can occur even in settings where people do not infer that popular alternatives are better. The model also casts new light on the effect of online recommendation systems on attitude homogenization and the effect of majority voting on beliefs and attitudes.

Keywords: Status Quo | Change | Learning | Social Influence | Popularity

*Both authors contributed equally to this work. Jerker Denrell: University of Warwick, Coventry, United Kingdom (Jerker.Denrell@wbs.ac.uk). Gaël Le Mens: Universitat Pompeu Fabra, Barcelona Graduate School of Economics and Barcelona School of Management, Barcelona, Spain (Gael.Le-Mens@upf.edu). We are grateful for discussions with and comments by Judith Avrahami, Manel Baucells, Jonathan Bendor, Nick Chater, Peter Glynn, Robin Hogarth, Yaakov Kareev, James Kitts, Balazs Kovaecs, Tomás Lejarraga, Jim March, Johannes Müller-Trede, Saurabh Prakash, Karl Schlag, Helena Szrek and Ezra Zuckerman. We benefited from comments by participants at the Annual meeting of the Academy of Management, 2008, and at the Annual Meeting of the American Sociological Association, 2008, the Barcelona Conference on Analytical Sociology and Social Mechanisms (GSADI), 2010, the INSEAD ICON seminar, 2010, the third TOM Meeting, Barcelona, 2012, Tilburg University, 2012, Chicago Booth School of Business, 2012, Universided Carlos III de Madrid, 2012; Stanford GSB, 2013. Gaël Le Mens benefited from financial support from the Barcelona School of Management, a Juan de la Cierva Fellowship from the Spanish Ministry of Science and Innovation and the Spanish Ministry of Science and Innovation Grants ECO2010-17145 and PSI2013-41909-P. Correspondence concerning this article should be addressed to Gaël Le Mens (Gael.Le-Mens@upf.edu).
1 Introduction

Explaining why collective mistakes emerge and persist is central to understanding social change, immobility, and differences in welfare (Bowles, 2004; Elster, 1978; North, 1990). In developing countries, people keep on using poor domestic hygiene practices even though simple changes would save many lives (Curtis, Cairncross & Yonli, 2000), and farmers fail to use fertilizer despite their potential for large increases in productivity (Duflo, Kremer & Robinson, 2009). In western countries, firms persist in making abundant use of temporary workforce despite the inefficiency of this arrangement (Pfeffer, 1998), and medical practices and treatments tend to cluster geographically, sometimes in favor of suboptimal alternatives (Coleman, Katz and Menzel, 1966; O’Connor et al., 1999; Krumholz et al., 2003). How come groups of people could persist in using inefficient technologies, practices or institutions when better courses of actions are available?

Prior literature has shown that such mistakes can occur when there are network externalities (Arthur, 1989; Elster, 1978; North, 1990; Pierson, 2004) or when a collective illusion emerges in which individuals mistakenly infer that the most popular but suboptimal alternative is the best (Banerjee, 1992; Bikhchandani, Hirshleifer & Welch, 1992; Cialdini & Goldstein, 2004; Festinger, 1957). In the first case, people are likely to herd, i.e. select the popular alternative, even if they know it is inferior to an unpopular alternative because it is difficult to coordinate a switch to the superior alternative. In the second case, people are assumed to make inferences about quality on the basis of popularity. More precisely, if a suboptimal alternative has, by chance, become the most popular one, people who have limited information about the qualities of the alternatives will mistakenly infer that the suboptimal but popular alternative is superior to other, less popular alternatives. This leads to herd behavior where most people chose the same alternative. Such inference can be rational, as in information cascade models (Banerjee, 1992; Bikhchandani el al., 1992), but it can also be the product of belief adjustments whereby people tend to process information about popular alternatives in a systematically different fashion from how they process information about unpopular alternatives (Bem, 1972; Elster, 1983; Festinger, 1957; Heider, 1958; Samuelson & Zeckhauser, 1988).

Here, we demonstrate theoretically that an alternative explanation for collective illusions is possible; an explanation that builds on the sampling perspective on belief formation (Denrell, 2005; Fiedler and Juslin, 2006). We consider settings where people learn about the qualities of the al-
ternatives from experience. In contrast to prior work, we assume that popularity does not impact inferences about qualities but only impacts sampling. In particular, we assume that due to conformity pressures or network externalities, people are more likely to sample alternatives the majority believes to be superior. We demonstrate theoretically that this mechanism, by itself, is sufficient to explain the emergence and persistence of collective illusions. Specifically, we show that if people tend to select popular alternatives, beliefs tend to become synchronized: a strong majority of people will tend to believe that one or the other alternative is the best, even if the qualities of the alternatives are the same. When a strong majority emerges in favor of the inferior alternative, a collective illusion has emerged. Our model shows that such a collective illusion is persistent in the sense that it tends to last for a long time.

Importantly, our theoretical results hold even if people do not engage in belief adjustments of the sort described above nor make inferences about quality on the basis of popularity. Our assumptions are thus very different from the assumptions of information cascade models. These models assume that people select the alternative that the majority of others has selected because of the inferences they make on the basis of popularity (Bikhchandani, Hirshleifer & Welch, 1992). Our model thus provides an existence proof that beliefs can become synchronized even in settings where people are reluctant to infer that what is popular is best for them, such as when they have different tastes or skills.

The intuition behind our results is that when popularity impacts sampling a population is unlikely to move away from a state in which one alternative is very unpopular compared to other alternatives. If by chance most people come to dislike alternative A, they are all unlikely to sample it again and their opinions of A remain the same. When A is in fact the best alternative, a collective illusion has emerged because people mistakenly believe that a suboptimal alternative is the best. Such a collective illusion is persistent because people avoid alternative A and do not correct their mistaken negative assessment of alternative A.

Our model builds upon prior work which has demonstrated how sampling can systematically impact belief formation when people learn from experience (Denrell, 2005; Denrell & Le Mens, 2007; Fiedler, 2000; Fazio, Eiser & Shook, 2004; Le Mens & Denrell, 2011). The key mechanism in much of this literature is the asymmetry in experiential learning generated by adaptive sampling. To ensure favorable future outcomes it makes sense to choose alternatives with favorable past outcomes and
avoid alternatives with unfavorable past outcomes. Such adaptive sampling implies that negative attitudes are more likely than positive attitudes to remain stable. If I dislike A, I am unlikely to sample A again and my negative attitude towards A does not change. Prior literature has used this mechanism to offer alternative explanations of a range of social psychological phenomena. For example, Denrell (2005) shows how in-group bias could be explained by adaptive sampling. The intuition is that attitudes will be more positive towards people you are likely to meet irrespective of your attitude towards them. If I dislike B, but B is a friend of my friends I might meet B again. As a result, my negative attitude towards B can change. This would not have happened if B were not part of my social network. They key mechanism is that attitudes will become more positive for alternatives the decision-maker is ‘forced’ to sample.

In this paper we consider a population of individuals who are all more likely to select alternatives that many others like. As in prior sampling models, sampling behavior is determined by outcomes in prior periods. Unlike prior models, however, there is no inbuilt bias towards sampling a particular set of alternatives (such as people belonging to the in-group). Rather, the sampling bias emerges endogenously (it is determined by what people sample and what outcomes they receive). We ask whether and when such an endogenously emerging sampling bias can generate a persistent collective illusion. Our model thus differs from prior models both conceptually (no inbuilt sampling bias in favor of a specific alternative) and in the focus on a collective (rather than individuals). Analyzing a model with an endogenously emerging sampling bias also requires mathematical techniques different from what was used in prior literature.

In the following, we first explain the rationale behind our assumption that people tend to select alternatives that many others like while not using popularity as a signal of quality nor engaging in belief adjustments. In Section 3, we analyze a simple learning model that gives rise to an information bias that leads to synchronized beliefs and to the emergence of collective illusions. In Section 4, we discuss some extensions to more realistic assumptions regarding choices and belief updating mechanisms. In Section 5 we discuss how our theory relates to prior research on the emergence of collective illusions. In Section 6, we show how our model can be adapted to explain how majority voting can lead to attitude homogenization within groups. Finally, in Section 7, we discuss the difficulties associated with empirically disentangling the sampling effect at the core of our approach and the effect of popularity on quality estimates that is mediated by popularity-based inferences.
2 Majority Influence and Information Sampling

Much prior research on why people might come to agree with a mistaken majority assumes that people change their beliefs to conform to the majority opinion. Theories of social learning in Economics and of majority influence in Psychology, for example, assume that people use popularity as an indication of quality: if most others like Alternative A better than Alternative B, people infer that Alternative A is likely superior (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992; Darley & Latané, 1968; Deutsch & Gerard, 1955; Prentice & Miller, 1993). Such inference about quality on the basis of popularity is often sensible: if most others like and go to restaurant A and few others like and go to restaurant B this suggests that A offers superior food and service. An abundant literature in Economics has shown that, at the population level, such inferences can lead to information cascades in which most agents come to mistakenly believe that a suboptimal alternative is optimal (a collective illusion has emerged). A key insight of this literature is that such information cascades, associated collective illusions and resulting herd behavior can be the outcome of rational behavior (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992; for a review, see Chamley, 2004).

While popularity-based inferences are often sensible, here we wish to demonstrate that such inferences are not necessary to explain the synchronization of opinions in a population. For this purpose, our model deliberately excludes such inferences. Specifically, we do not assume that people change their beliefs when hearing about the behaviors of others: in our model, people’s quality estimates are based only on the information they obtain by sampling the alternatives. We assume, however, that people are subject to majority influence in the sense that they are more likely to select and thus sample an activity that is popular, i.e., believed to be of superior quality by most other people. Thus, our model assumes majority influence but it does not assume a direct effect of popularity on beliefs (beliefs depend only on one’s information sample).

Why assume that the majority only impacts sampling behavior and not beliefs? In cases where popularity-based inferences are important, this assumption may not be realistic since popularity then likely impacts both behavior and beliefs. For theoretical purposes, however, it is useful to examine
the case where popularity only influences behavior. We will show that this assumption, combined
with individual learning from experience, is sufficient to explain emergence of belief synchronization.
This strategy that consists in assuming away intrapsychic belief adjustments is standard in the
research tradition in which our paper fits (Fiedler & Juslin, 2006; Juslin, Winman, & Hansson, 2007).
Like other information sampling model, our theory emphasizes the role of the environment and how
it can sometimes produce systematically biased samples of information that affect judgments (see
Fiedler & Juslin, 2006 for a review and Denrell & Le Mens, 2011; Eder, Fiedler & Hamm-Eder,
Mens & Denrell, 2011; Smith & Collins, 2009, for recent publications that adopt this perspective).

There are also many realistic situations where one might expect to see the majority impact
sampling behavior whereas people do not engage in popularity-based inferences. Prior studies have
shown that people might publicly adhere to the local norm, by stating an opinion in agreement with
the majority or choosing the action chosen by most others, while privately disagreeing or having
different preferences. In the Asch conformity experiments (1956), for example, many participants
did not believe that the majority was correct but nevertheless stated an opinion in agreement with
the majority opinion to avoid being seen as deviants. More generally, studies have shown that
people who deviate from the majority opinion are less liked (Levine, 1980) and also expect to be
disliked (Gerard and Rotter, 1961), so it is hardly surprising that people conform publicly even
if they privately are of a different opinion. People may also go along with the majority because
there is ‘safety in numbers’ in the sense that it can be costly to choose an activity few others will
choose (Granovetter, 1978; Schelling, 1978). For example, a research student will find it easier to
get help if she chooses a research method used by many faculty and students in her department.
The advantage of having help around may motivate her to choose this method even if she does
not believe that it is overall superior. People might also choose the popular alternative because
it is “better to fail conventionally than to succeed unconventionally” (Keynes, 1936, p. 158), i.e.,
because of the adverse reputation effects of receiving a poor outcome with an unusual alternative
(Scharfstein & Stein, 1990; Zwiebel, 1995). Finally, people might select popular alternatives because
they are more easily available or visible. For example, people are more likely to be aware of popular
songs because they are played more often on radio etc. Popular alternatives are often also more
prominently displayed on websites (such as Google, Spotify etc).
Besides, while popularity-based inferences are often sensible, there are settings in which it is doubtful whether the majority knows best or knows what is good for you. People may have unique tastes (Berger & Heath, 2007): if others prefer opera to jazz, this may not be enough to convince you that opera is superior. But you might nevertheless go to the opera because you want to be able to talk about it with your friends. Individuals may also suspect that the experiences of others are not relevant because their circumstances are different. For example, because of wide variations in soil conditions, rice growers during the Indian green revolution were hesitant to adopt techniques simply because they had worked well at neighboring plots (Munshi, 2004). Rather than imitating others, they relied on their own experience. Such hesitance to rely on the experience of others is likely common and also sensible especially when the experiences of others contradict your own experience: the fact that the experiences of others differ from yours indicates that your circumstances may differ substantially. The possibility of different circumstances was also one of the explanations that participants in the Asch experiments (1956) came up with to make sense of the behavior of the majority: the fact that the majority disagreed with them indicated that the majority interpreted the question differently or viewed the lines from a different angle. Individuals may also rely more than they should on their own experience simply because it is salient (Simonsohn et al, 2008).

These arguments suggest that there are settings in which popularity-based inferences are unlikely to be important but people might nevertheless select the popular alternative to avoid being seen as deviant, to coordinate their choices with others, or because popular alternatives are more easily available. In these cases popularity will influence sampling but not beliefs once sampling is controlled for. A good illustration is the MusicLab experiment (Salganik et al., 2006). In this experiment, thousands of participants could listen to songs by streaming them and then decide to download (or not) the songs to which they had listened. Songs that had been downloaded by many participants (i.e. the popular songs) were more accessible in that they were displayed at the top of the computer screen (Experiment 2). Recent re-analyses of the experimental data suggest that participants made their listening (i.e. sampling) decisions based on popularity. But their evaluations of the songs were not driven by social influence: there was no effect of popularity once sampling was controlled for (Hendricks, Sorensen & Wiseman, 2012; Krumme et al., 2012).
3 Model

We consider a population of \( n \) people who learn about the qualities of two uncertain alternatives from the private information they obtain by sampling the alternatives. In each period, one individual is randomly picked in the population. This individual selects one of the two alternatives, observes its payoff and updates her estimate of the quality of this alternative accordingly. We implement the effect of network externalities by assuming that the individual is more likely to sample the alternative that is believed by most to be the superior one. We focus on the essential features of the learning and choice processes that can give rise to collective illusions. Therefore, we keep our model as simple as possible in this section. In section 4 we discuss the assumptions of the model and what happens when we make different assumptions.

3.1 Model Specification

Our model is based on the following assumptions. Unless otherwise noted, we denote random variables by capital letters, and their instantiations by corresponding lower-case letters. Vectors are referred to using bold font.

**Payoffs of the Alternatives** The two alternatives potentially differ in ‘quality’ or ‘efficiency’. The two alternatives have payoffs that follow Bernoulli distributions: The possible payoffs are 0 (failure) and 1 (success). Let \( p_1 \) be the probability of success of Alt. 1 and \( p_2 \) the probability of success of Alt. 2. We will say that Alt. 1 is superior (inferior) when \( p_1 > p_2 \) (\( p_1 < p_2 \)). An important feature of the model is that the success probabilities, \( p_1 \) and \( p_2 \), are unknown to the decision maker.

**Quality Estimates** Estimates of the qualities of the alternatives are binary: an individual believes an alternative is either ‘good’ or ‘bad’ (and so she can believe that both are ‘good’, both are ‘bad’ or one is ‘good’ while the other is ‘bad’). The estimate of Alt. 1 (Alt. 2) for Individual \( i \), at the beginning of period \( t \) will be denoted \( \hat{Q}_{1,i,t} \) (\( \hat{Q}_{2,i,t} \)). If the estimate of Alt. 1 (Alt. 2) is ‘good’, we will write \( \hat{Q}_{1,i,t} = 1 \) (\( \hat{Q}_{2,i,t} = 1 \)) and if the estimate is ‘bad’, we will write \( \hat{Q}_{1,i,t} = 0 \) (\( \hat{Q}_{2,i,t} = 0 \)). Initial estimates are random draws from the payoff distributions. More precisely, for a given individual, the likelihood that the initial estimate of Alt. 1 is ‘good’ is \( p_1 \). The likelihood that the initial estimate
is ‘bad’ is $1 - p_1$. If $p_1 = .5$, then an individual initially believes that Alt. 1 is ‘good’ or ‘bad’ with equal probabilities.

**Sampling Rule** In each period, one individual is randomly picked in the population (each individual is equally likely to be picked). This individual makes a choice between the two available alternatives. The crucial assumption is the following: we assume that the individual is more likely to sample the popular alternative than the unpopular alternative. Specifically, let $N_{1,t}$ be the number of individuals who have a ‘good’ estimate for Alt. 1 at the beginning of period $t$ and $N_{2,t}$ the number of individuals who have a ‘good’ estimate for Alt. 2 at the beginning of period $t$. We assume that the likelihood that an individual chooses the first alternative is

$$p_{C_1}(n_{1,t}, n_{2,t}) = \frac{1}{1 + e^{-s(n_{1,t} - n_{2,t})}},$$

where $s > 0$ is the sensitivity to popularity. We set $p_{C_2}(n_{1,t}, n_{2,t}) = 1 - p_{C_1}(n_{1,t}, n_{2,t})$. This logistic choice rule is often used to model choice under uncertainty (Erev & Barron, 2005, Sutton & Barto, 1998) and provides good fit to experimental data on sequential choices (Busemeyer & Myung, 1993; Denrell, 2005). It also fits data on the effect of the number of influence sources on conformity behavior (MacCoun, 2012).

**Estimate Updating Rule** We assume that quality estimates are based on the most recent observation. If Individual $i$ samples Alt. $k$ and gets a success, her estimate becomes ‘good’ (e.g. for Alt. 1, $\hat{Q}_{1,i,t} = 1$). If she gets a failure, her estimate becomes ‘bad’ (e.g. for Alt. 1, $\hat{Q}_{1,i,t} = 0$). We make this assumption to keep the model as transparent as possible, but the basic results continue to hold if individuals use a weighted average estimate updating rule, as shown in Section 4.1. Although this assumption that only the most recent observation matters might seem simplistic, recent experimental evidence on sequential choice between uncertain alternatives with binary outcomes shows that choices are well predicted by a model that focuses on the most recent outcome only, like ours (Avrahami & Kareev, 2010).

---

1 Unless otherwise noted, capital letters refer to random variables. Corresponding non-capital letters denote specific instances of the random variables. We use non-capital letters to refer to non-random model parameters.

2 See Section 4.2 for further discussion of the logistic choice rule.
Figure 1: Features of the stationary distribution. (a) The stationary joint distribution of the popularities of Alt. 1 and Alt. 2 is bimodal. (b) The marginal distribution of $N_1 - N_2$ is also bimodal. Graphs obtained with $n = 10, p_1 = 0.40, p_2 = 0.45, s = 2$.

3.2 Results

In what follows, we first present an analysis of the asymptotic behavior of the model because such an analysis leads to easily interpretable formulas that make it easier to understand the pattern of behavior implied by the model. Then we demonstrate, using numerical computations, that our main results emerge after a reasonable number of periods.

3.2.1 Asymptotic Analysis

To demonstrate the synchronization of quality estimates, we computed the stationary distribution of the popularities of the two alternatives. Let $N_1$ and $N_2$ denote the random variables toward which $N_{1,t}$ and $N_{2,t}$ converge and let $\bar{f}_t(n_1, n_2)$ denote the joint stationary distribution of the popularities of $N_1$ and $N_2$.

**Proposition 1.** [Joint Stationary Distribution of Popularities] When sampling follows the logistic choice rule (eq. 1), the stationary distribution of the popularities of the two alternatives, $N_1$ and $N_2$, is

$$\bar{f}(n_1, n_2) = K \left( e^{-sn_1} + e^{-sn_2} \right) (p_1)^{n_1} (1 - p_1)^{n-n_1} (p_2)^{n_2} (1 - p_2)^{n-n_2} \binom{n}{n_1} \binom{n}{n_2}, \quad (2)$$

where $K^{-1} = (1 - p_1 + p_1 e^{-s})^n + (1 - p_2 + p_2 e^{-s})^n$.

**Proof.** The proofs are in the Appendix. \qed
This distribution implies that a strong majority favoring one or the other alternatives is likely to emerge (see Figure 1a). To understand why, note that this is a weighted bivariate binomial distribution, where the weight is $K(e^{-sn_1} + e^{-sn_2})$. The weighting function puts additional weights on the extremes (where $N_k = 0$) and implies that the joint distribution of popularities will be bi-modal if $s$ is high enough. The first mode is close to $(N_1 = p_1n, N_2 = 0)$: in this case, the proportion of people who like Alt. 1 is $p_1$ while the proportion of those who like Alt. 2 is small. The other mode is located at $(N_1 = 0, N_2 = p_2n)$: in this case the proportion of people who like Alt. 1 is small while the proportion who like Alt. 2 is $p_2$.

Figure 1b shows the distribution of the difference in popularities, $N_1 - N_2$, which is also bimodal. One mode is above 0; this corresponds to cases where Alt. 1 is the more popular. The other mode is below 0; this corresponds to cases where Alt. 1 is the less popular. In the setting of the graph, Alt. 1 is inferior to Alt. 2 ($p_1 = 0.4$ and $p_2 = 0.45$). Unsurprisingly, the mode above 0 is lower than the mode below 0; Alt. 1 is less likely than Alt. 2 to become the more popular. But the probability that Alt. 1 becomes more popular than Alt. 2 is still substantial (the probability is 0.29) despite the fact that Alt. 1 is the inferior alternative. When Alt. 1 is more popular despite being inferior, a collective illusion has emerged: most people mistakenly believe Alt. 1 to be superior to Alt. 2. We formalize the definition of collective illusion as follows:

**Definition.** [Collective Illusion] We will say that the system is in a collective illusion state when either of the following two situations arise:

- $p_1 < p_2$ and $N_1 > N_2$
- $p_1 > p_2$ and $N_1 < N_2$

The intuition for Proposition 1 is that when popularity impacts sampling the system is unlikely to move away from a state in which one alternative is unpopular. This is related to the bimodal character of the asymptotic joint distribution of popularities. To explain this, suppose $p_1 < p_2$ and $N_1 = np_1$ while $N_2$ is low (i.e. $N_1 > N_2$). Thus, Alt. 1 is the most popular and Alt. 2 is believed by many to be ‘bad’. The system can only exit from such a collective illusion if $N_2$ increases or $N_1$ decreases. These events are unlikely to happen, however.

---

3With the parameters used for constructing the graphs of Figure 1, the first mode is located at (4,1) and the second mode is located at (0,4).
To see why, note first that when Alt. 2 is unpopular, it is unlikely to be selected. So, \( N_2 \) is unlikely to increase. Next, note that while Alt. 1 is likely to be selected (because it is popular), \( N_1 \) is unlikely to decrease to a value much below \( np_1 \). When Alt. 1 is selected, the probability that \( N_1 \) decreases by one is \( \frac{N_1}{n} (1 - p_1) \) while the probability that \( N_1 \) increases by one is \( (1 - \frac{N_1}{n})p_1 \). An increase is more likely than a decrease whenever \( N_1 < np_1 \), a decrease is more likely if \( N_1 > np_1 \), and a decrease is as likely as an increase when \( N_1 = np_1 \). In other words, when Alt. 1 is selected, \( N_1 \) tends to regress to its mean, which equals \( np_1 \). Overall, this implies that system the is very slow to move away from the collective illusion in which \( N_1 = np_1 \) and \( N_2 \) is low.

This reasoning also shows that a collective illusion is a not temporary chance occurrence that will be quickly corrected. Rather, collective illusions produced by our model tend to be self-sustaining: once a collective illusion has emerged, the system is very slow to exit this state. To illustrate the persistence of collective illusions we computed the expected time the system stays in the collective illusion state once a collective illusion has emerged. In what follows, we assume that Alt. 2 is superior to Alt. 1 (\( p_1 < p_2 \)).

---

\[ \text{Formally, suppose that we are at the beginning of period } t_0. \text{ Suppose that } t_0 \text{ is large such that we can assume that } N_{1,t_0} \text{ and } N_{2,t_0} \text{ follow the asymptotic joint distribution of popularities given by eq. 2. Let } T \text{ denote the first time the system enters a state such that } N_{1,T} > N_{2,T}. \text{ Using numerical computations, we compute the expected 'hitting' time of the region of the state space where } N_1 > N_2. \text{ This is } E[T - t_0|N_{1,t_0} < N_{2,t_0}]. \]
Figure 2a reports the results of these computations. The graph displays the expected number periods the collective illusion persists, given that we know that there is a collective illusion, as a function of the slope parameter of the logistic sampling rule \((s)\). The graph shows that collective illusions tend to persist longer when sampling strongly depends on popularity \((s\) is high). It is useful to note that even when choices are independent \((s = 0)\), it still takes some time for the system to exit the collective illusion (because only one individual can sample and change beliefs in each period). But the crucial feature of this figure is that collective illusions persist much longer when people tend to sample popular alternatives.

Figure 2b shows how this pattern changes with the difference between the qualities of the two alternatives. Intuition suggests that when the alternatives have very different qualities, collective illusions are unlikely to persist. The reason is that, in this case, it is very easy for the agents to learn from their experience that the alternatives are different. We find that this is the case. The graphs of Figure 2b are plotted by assuming that Alt. 1 is inferior to Alt. 2. More precisely, \(p_1 = 0.5 - \beta\) and \(p_2 = 0.5 + \beta\), where \(\beta\) is a positive parameter that regulates the difference between the qualities of the alternatives. When \(\beta\) is small, the two alternatives have very similar qualities, and when \(\beta\) is larger, Alt. 2 is much better than Alt. 1. Figure 2b shows that when \(\beta\) is small, collective illusions are persistent. But when \(\beta\) is large, collective illusions do not persist, even if people have a strong tendency to select the popular alternative (i.e. \(s\) is high).

We have just seen that the difference between the qualities affects the persistence of collective illusions. But what about the likelihood of emergence of collective illusions? Collective illusions are less likely to emerge when the alternatives have very different qualities and more likely to emerge when the qualities are similar (see Figure 3 ). The intuition for this result is again that is easier for agents to learn which alternative is the best when the qualities are very different because each observation is more diagnostic than when the alternatives have similar qualities.

Overall, this discussion implies that collective illusions are more likely to emerge and persist when the difference between the qualities of the alternatives is small. Our model shares this feature with some of the existing models that predict the emergence and persistence of collective illusions. In particular, the graph of Figure 3 is very similar to Figure 1 in Bikhchandani et al. (1992), which plots the likelihood of an incorrect cascade as a function of the accuracy of the signal received by the agents. Our model, however, relies on a different mechanism and makes predictions that are
3.2.2 Dynamics of the Model: Numerical Computations

So far, we discussed properties of the asymptotic distribution of popularities because it has an easily interpretable formulation as a weighted bimodal distribution. Although it does not seem possible to derive a similarly interpretable formula for the distribution of popularities in any given period, it is easy to compute the distribution numerically (by computing successive powers of the transition matrix implied by the above assumptions). Figure 4a shows the joint distribution of popularities at the beginning of period 100, assuming that all agents in the model start with independent beliefs. The joint distribution of popularities is similar to the asymptotic distribution of Figure 1a. It is bimodal. One mode is such that $N_1$ is close to $np_1$ and $N_2$ is low. The other mode is such that $N_1$ is low and $N_2$ is close to $np_2$.

We can illustrate how the effect unfolds over time by plotting the distribution of $N_{1,t} - N_{2,t}$ for various values of $t$. Figure 4b shows that the distribution of the difference is initially unimodal, and it becomes bimodal after a number of periods somewhere between 50 and 100. This means that after about 10 decisions per agent, if a collective illusion emerges, it will have a self-sustaining character. The self-sustaining character of the collective illusion becomes stronger as the distribution becomes
more strongly bimodal. The bimodal character of the distribution becomes more pronounced as the number of period grows. After about 1000 periods, the distribution has essentially reached its asymptotic state.

3.2.3 Positive Dependence Between Quality Estimates

The synchronization of estimates across individuals is most clearly illustrated by analyzing the correlation between quality estimates of pairs of individuals. The model implies that quality estimates for a given alternative become positively correlated. This occurs despite the fact that the individuals in the model make independent observations of the payoffs of the alternatives. Conversely, the quality estimates for competing alternatives become negatively correlated.

Proposition 2. [Correlation Between Quality Estimates of Two Individuals]

(i) The pairwise correlations between stationary quality estimates for a given alternative are

---

5These results are consistent with prior work on the emergence of correlated estimates and attitudes in a dyad as a result of interdependent sampling (Denrell & Le Mens, 2007). In fact, Proposition 2 extends Theorem 1 in Denrell & Le Mens (2007) and shows that similar results apply to settings where there are more than two individuals and one alternative and where people do not sample simultaneously.
Figure 5: Odds ratios that a given individual believes an alternative is ‘good’ as opposed to ‘bad’ when another individual believes that the same alternative is ‘good’ or ‘bad.’ (a) Alt. 1. (b) Alt. 2. Graphs obtained on the basis of 10,000 simulations with \( n = 10, p_1 = 0.4, p_2 = 0.45, s = 2 \).

(ii) The pairwise correlations between stationary quality estimates for competing alternatives are negative. For all \( i \neq j \),

\[
\text{corr} \left( \hat{Q}_i^1, \hat{Q}_j^2 \right) < 0.
\]

Although these results about correlation between quality estimates are elegantly formulated, our assumption that the payoff distribution is binary implies that correlations are almost surely very low. As such, the correlations do not provide a good intuitive sense about the size of the effect.

The following corollary provides an alternate formulation of the same idea in terms of odds ratios.

**Corollary 1.** [Dependency between Quality Estimates] The odds that a given individual believes an alternative is ‘good’ as opposed to ‘bad’ is higher when another individual believes so as well. For all \( i \neq j \), and for \( k \in \{1, 2\} \),

\[
\frac{P(\hat{Q}_k^i = 1|\hat{Q}_k^j = 1)}{P(\hat{Q}_k^i = 0|\hat{Q}_k^j = 1)} > \frac{P(\hat{Q}_k^i = 1|\hat{Q}_k^j = 0)}{P(\hat{Q}_k^i = 0|\hat{Q}_k^j = 0)}.
\]

Figure 5 displays how these odds ratios evolve over time. The inequality predicted by Corollary
1 emerges rather quickly, after about 50 periods, and increases over time. After about 50 periods the odds that an individual $i$ believes that Alt. 1 is ‘good’ rather than ‘bad’ when another individual $j$ also believes Alt.1 is ‘good’ are higher than the odds that $i$ believes that Alt.1 is ‘good’ rather than ‘bad’ when $j$ believes Alt. 1 is ‘bad’ (see left quadrant). The same holds for Alternative 2 (see right quadrant). In summary, the quality estimates of the individuals in the group become positively dependent on each other after some time.

These results about the emergence of a positive dependence between estimates are important because they illustrate a possible limitation of the wisdom of crowds (Galton, 1907; Surowiecki, 2004). The wisdom of crowds refers to the observation that the belief of the majority is often correct. The theoretical explanation of this phenomenon relies on the Condorcet jury theorem (Condorcet, 1785) which assumes that beliefs are formed independently (Grofman, Owen, & Feld, 1983). It is well-known that if individuals base their beliefs on the beliefs of others, i.e., if social learning operates, then the majority beliefs may not contain much more information than the belief of a single individual (e.g. Lorenz et al., 2011). This is because the beliefs of the members of the group are based on redundant information. Our theory implies that even if beliefs are based on independent pieces of evidence, beliefs will still become dependent if sampling behavior is affected by popularity.

### 3.3 Relation to Prior Sampling Models

Our formal model is based on but considerably extends prior sampling models. As in prior sampling models (e.g., Denrell, 2005) we examine how beliefs evolve when prior beliefs influence sampling. Unlike prior models, we consider a population of individuals, who make their own sampling decision although influenced by majority beliefs. The most closely related prior model is the model of interdependent sampling by Denrell and Le Mens (2007). Like the current model, the model of interdependent sampling examines how the attitudes of several individuals evolve over time when they influence each others’ sampling processes. Unlike the current model, however, the model of interdependent sampling assumes that individuals sample simultaneously (for example, two individuals go together to the same restaurant). This assumption can be realistic for dyads but is unrealistic for large populations. Our current model examines a population of individuals who sample alternatives at different times but whose sampling behavior is influenced by the opinions of others. The
models thus differ both conceptually and formally. Moreover, while the focus of Denrell and Le Mens (2007) was on when joint sampling could lead to positive or negatively correlated attitudes, this paper focuses on how a majority influence on sampling behavior can lead to the emergence of a collective illusion.

4 Discussion of the Assumptions of the Model

Our basic results also hold when there are more than two alternatives and under different assumptions about the payoff distributions and updating rule. The payoff distributions could be Normal instead of Bernoulli and the quality estimates could be weighted averages of all past observations instead of being based only on the most recent observation. In this section, we provide a complete description of this version of our model because it leads to formulas that allow for an easy interpretation of the effects of the various model parameters. We also discuss other influences on sampling and belief updating.

4.1 A model with Normally distributed payoffs and fractional updating

4.1.1 Model Specification

The structure of the model is the same as before, but the assumptions regarding the payoff distributions and estimate updating rules differ.6

Payoffs of the Alternatives The payoffs of Alt. $k$ ($k \in \{1, 2\}$) are independent draws from a normal distribution with mean $\mu_k$ and variance $\sigma^2$ (both alternatives have the same variance). In what follows, $\mu_k$ refers to the ‘quality’ of Alternative $k$.

Quality Estimates Individual $i$’s quality estimate for Alt. 1 (Alt. 2) at the beginning of period $t$ is random variable $\hat{Q}_{1,t}^i$ ($\hat{Q}_{2,t}^i$). The initial estimate of Individual $i$ for the quality of Alt. $k$ is a random draw from a normal distribution with mean $\mu_k$ and variance $\sigma^2$. Let, for all $i$, $\hat{Z}_t^i = (\hat{Q}_{1,t}^i, \hat{Q}_{2,t}^i)$ and $\hat{Z}_t = (\hat{Z}_1^1, ..., \hat{Z}_n^p)$. Similarly, for all $i$, $\hat{z}_t^i = (\hat{q}_{1,t}^i, \hat{q}_{2,t}^i)$ and $\hat{z}_t = (\hat{z}_1^1, ..., \hat{z}_n^p)$.

---

6For ease of exposition, our model includes just two alternatives. But our results are easily adapted to the case where there are $M > 2$ alternatives.
**Sampling Rule**  The probability that an individual selects the first alternative is given by a logistic choice rule, as in the model of the previous section. The choice likelihood is based on the average quality estimates of the two alternatives, computed across all the people in the population:

\[
 p_{C_1,t} = p_{C_1}(\hat{z}_t) = \frac{e^{\frac{1}{n} \sum_{i=1}^{n} \hat{q}_{1,i,t}}}{e^{\frac{1}{n} \sum_{i=1}^{n} \hat{q}_{1,i,t}} + e^{\frac{1}{n} \sum_{i=1}^{n} \hat{q}_{2,i,t}}},
\]

where \( s \) is a parameter that regulates the sensitivity of the choice probability to the average quality estimate. We assume \( s > 0 \). This equation implies that the individual is likely to select the more popular alternative, in the sense that people in the population estimate it more positively. The probability of selecting the second alternative is defined similarly and is denoted by \( p_{C_2,t} = p_{C_2}(\hat{z}_t) = 1 - p_{C_1,t} \).

**Estimate Updating Rule**  The revised estimate is a weighted average of the past estimate and the new observation (Busemeyer & Myung, 1992; Denrell, 2005). When Individual \( i \) selects Alt. 1, her quality estimate is updated as follows:

\[
 \hat{Q}_{i,1,t+1} = \hat{Q}_{i,1,t}(1 - b) + bO_{i,1,t},
\]

where \( O_{i,1,t} \) is the observation of the payoff of Alt. 1 in period \( t \) \((O_{i,1,t} \sim N(\mu_1, \sigma^2))\) and \( b \) is such that \( 0 < b < 1 \). When an alternative is not sampled, its quality estimate does not change. The estimate updating rule is the same for Alt. 2.

**4.1.2 Analysis**

In this setting, it is possible to derive easily interpretable closed form expressions for the correlations between the quality estimates of any two individuals.\(^7\) We denote by \( \hat{Q}_1 \) (\( \hat{Q}_2 \)) Individual \( i \)'s stationary quality estimate of the first (second) alternative.

**Proposition 3.** *(Correlation Between Quality Estimates of Two Individuals)*

(i) The pairwise correlations between stationary quality estimates for a given alternative are positive. For all \( i \neq j \), and for \( k \in \{1, 2\} \),

\[
 \text{corr} \left( \hat{Q}_k, \hat{Q}'_k \right) = \frac{1}{1 + \left( 2 + e^{-s(\mu_1 - \mu_2)} + e^{s(\mu_1 - \mu_2)} \right) \frac{n^2(2-b)}{s^2\sigma^2b}} > 0. \tag{4}
\]

\(^7\)In this case, the joint distribution of the popularities has a form very similar to the joint distribution of popularities in the model analyzed above. See Lemma 3 in Appendix B for details.
(ii) The pairwise correlations between stationary quality estimates for competing alternatives are negative. For all \( i \neq j \),

\[
corr \left( \hat{Q}_i^1, \hat{Q}_j^2 \right) = -corr \left( \hat{Q}_k^i, \hat{Q}_k^j \right) < 0.
\]

The sizes of the correlation increases with the weight of new observations \( b \), the standard deviation of the payoff distribution \( \sigma \), and \( s \), the sensitivity of the choice likelihood to quality estimates. When \( b \) is close to 1, the effect is stronger because quality estimates reflect few observations (past observations are heavily discounted). With \( n = 5, \mu_1 = -0.25, \mu_2 = 0, \sigma = 2, s = 5 \), the correlation goes from 0 with \( b = 0 \) to 0.41 with \( b = 1 \). It is equal to 0.15 when \( b = 0.4 \), which is a typical value for \( b \) in several experiments on sequential choice under uncertainty (see Table 1 in Denrell, 2005). If \( \sigma \) is high, each observation is a ‘noisy’ signal of the quality of the alternative. The correlation goes from 0 with \( \sigma = 0 \) to 1 when \( \sigma \) is very large.

The correlation is maximal when the two alternatives have the same quality \( (\mu_1 = \mu_2) \) and it decreases with the difference between the qualities of the alternatives \( (|\mu_1 - \mu_2|) \). This is not surprising, because when the qualities of the alternatives are very different, each observation is more diagnostic of the ranking of the alternatives. Finally, the correlation decreases with \( n \), the size of the population. The intuition is that when the population is larger, the influence of each member on the sampling behavior of the focal individual is diluted. The reason is that the focal individual’s sampling likelihood is assumed to increase with the average estimate of all the members in the population.

### 4.2 The Sampling Rule and the Role of Popularity

Our assumption of a logistic sampling rule (see eq. 1 and eq. 3) is consistent with much prior research on sequential choice under uncertainty. It has been used in models with a reinforcement learning structure (e.g. Denrell & Le Mens, 2011; Erev & Barron, 2005; Sutton & Barto, 1998) and it fits well data on repeated choices between uncertain alternatives (e.g. Busemeyer & Stout, 2002). A version of the logistic choice rule has also recently been proposed by MacCoun (2012) as a model for the effect of the number of influence sources on conformity behavior. MacCoun (2012) shows that the model fits well data from several sources including data from Asch’s experiment on line length judgments (Asch, 1956) and the Milgram et al. (1969) study on the effect of the number of
people looking up at a 6th floor window in a street of New York city.

In the logistic choice model, the parameter $s$ regulates how sensitive people are to the opinions of others. A high value of $s$ implies that people are very likely to choose the more popular alternative. As MacCoun (2012) suggests, this parameter can be interpreted as an index of norm clarity. The magnitude of our results - the extent to which quality estimates are influenced by popularity - depends critically on the value of $s$. The effect is large when $s$ is high and small when $s$ is low. Thus, according to our model popularity cannot influence quality estimates much if choices are relatively insensitive to popularity (eq. 4 shows that the correlation between estimates increases with $s$). The same holds for many other models of the effect of popularity and conformity on belief formation, such as social impact theory (Latané, 1981; Latané & Wolf, 1981).

Our model assumes that all individuals in the population, including the focal actor, matter equally. Formally, sampling only depends on $N_1$ and $N_2$. Hence an individual is not assumed to weight his or her own opinion more than that of others. This assumption was made for simplicity and mathematical tractability (assuming symmetry seems necessary to derive an analytically tractable formula for the joint distribution of the estimates). Nevertheless, computer simulations show that our results still hold when this assumption is relaxed. For example if the weight of the estimate of the self is equal to 2, whereas the weight of the estimates of others is equal to .5, the joint distribution of popularities is still bimodal after 100 periods.\(^8\)

Our model assumes that people are more likely to choose popular alternatives but this assumption is not always correct. If an alternative has become popular individuals may wish to choose a less popular alternative to restore their freedom of choice (Brehm, 1966) or to distinguish themselves from others (Brewer, 1991). In particular, Brewer’s theory of optimal distinctiveness suggests that individuals shy away from alternatives that are too popular. What are the implication of such alternative assumptions for our results? On the one hand, if the preference for distinctiveness is strong, our basic result is unlikely to hold. This suggests an important scope condition for our model (as well as for other models, including herding models, that assume that people choose popular alternatives). On the other hand, our basic result can still hold if people care about distinctiveness in addition to popularity.

To verify this, consider the following version of our model with binary payoffs. Suppose that

\(^8\)For the model with binary payoffs, with $n = 10, p_1 = 0.4, p_2 = 0.45$ and $s = 2$.\(^8\)
the attractiveness of Alternative 1 increases linearly with the number of others who have a ‘good’
estimate of that alternative until $N_1 = \delta$ and then decreases linearly until 0 when $N_1 = n$. The same
is assumed for Alternative 2. Formally, we assume suppose

$$p_{C1}(n_{1,t}, n_{2,t}) = \frac{\alpha_g(n_{1,t})}{\alpha_g(n_{1,t}) + \alpha_d(n_{2,t})} \text{ with } \alpha_d(m) = e^{s(m \leq \delta + \frac{n - m}{N - m} m > \delta)}$$

and $I_{m \leq \delta}$ is an indicator variable equal to 1 when $m \leq \delta$.

Simulations of this model (with the same parameters as for Figure 1) show that when $\delta$ is low (e.g. $\delta = 2$), people avoid popular alternatives, and the distribution of the difference in popularity is unimodal with its mode close to 0 (after 100 periods). That is, when people have a strong preference for distinctiveness ($\delta = 2$), our main result does not hold. But if the preference for distinctiveness is weaker ($\delta$ is equal to 4 or higher), the distribution of the differences in popularity is bimodal: our main result holds in this case.

### 4.3 Belief Updating and Popularity

Our model assumes that popularity only influences sampling and has no direct effect on belief updating. In reality, popularity can impact both sampling and belief updating. Most importantly, people may infer that the popular alternative is better. We examine this possibility in the next section when we compare our model to models in the information cascade tradition (see section 5.3). Several other interactions between popularity and belief updating are possible. First, if people feel pressured to select the popular alternative - and experience this as a threat to their freedom - they might discount or ignore the information sampled under such pressure (Brehm, 1966). Second, those who consciously select an unpopular alternative - and thus demonstrate agency - may be motivated to assess it positively (Aronson, 2007). Experiences with the unpopular alternative may for this reason be interpreted differently than experiences with the popular alternative. Finally, confirmation bias and contrast effects imply that prior expectations impact how new experiences are interpreted. When going to a popular restaurant, people may have positive expectations and pay attention to experiences confirming their expectations (Nickerson, 1998). On the other hand, a negative experience with an alternative for which the decision maker had positive expectations can be particularly striking, and might carry a high weight in the belief updating process. In settings where any of these effects has a large impact on belief updating, our results may not hold. This suggests that our model is most applicable to settings where the influence of popularity is somewhat subtle, such as when people are nudged to select more popular alternatives because these are more
5 Relations to Other Explanations

5.1 Coordination Failures Due to Network Externalities

Our theory differs from explanations of collective mistakes that attribute them to a coordination failure due to network externalities (Arthur, 1989; Elster, 1978; North, 1990; Pierson, 2004). These explanations assume that people know that they are selecting a suboptimal alternative, but do not switch because it is difficult (or impossible) to coordinate the transition to the superior alternative. Our theory suggests that if network externalities lead people to select and thus sample popular alternatives, most people may not only choose the inferior alternative but, when this happens, most people will also come to believe that the inferior alternative is of superior quality. Thus, in a vote about whether to switch to another alternative, most people would favor sticking to the status quo even if it is actually inferior. By contrast, explanations that rely on a coordination failure predict that people will switch if a vote could be organized and switching costs were low.

5.2 Belief adjustments

Our theory also differs from existing mechanisms that explain why people might agree with the beliefs of the majority even when the majority is wrong. Existing explanations for this synchronization of beliefs across members of a social group generally rely on belief adjustments (Cialdini & Goldstein, 2004; Kunda, 1990; Wood, 2000). One class of explanations proposes that people might adopt the behavior of the majority and then adjust their (private) beliefs and attitudes with their own behavior (Bem, 1972; Festinger, 1957). Another class of explanations proposes that people might adjust their (private) beliefs and attitudes to conform with the behavior of the majority because they identify with the majority (Cialdini & Goldstein, 2004; Festinger, Schachter & Back, 1950; Hogg & Turner, 1987; Turner, 1991; Wood, 2000). Our explanation does not challenge the experimental evidence underlying those explanations. Rather, it suggests a complementary explanation that operates at a different level of analysis: these explanations invoke properties of the mind’s information processing mechanisms whereas our explanation focuses on properties of the sample of information on which these mechanisms operate. In order to analyze the role of information available.
sampling biases, we have assumed that people are good processors of information but that they are naïve with respect to the nature of the sample they use to form their beliefs. This strategy is central to the research program on sampling explanations of judgment biases in which our theory fits (Fiedler 2012; Fiedler & Juslin 2006; Juslin, Winman & Hansson, 2007). Just like other contributors to this research program, we do not claim that people do not engage in such intrapsychic belief adjustments. Empirically analyzing how such belief adjustments interact with information sampling mechanisms is a necessary next step in this research program.

5.3 Information Cascades

How do our results relate to herding models that rely on information cascades, which also explain collective failures to identify the best alternative (Banerjee, 1992; Bikhchandani et al., 1992)? At an abstract level our model may seem quite similar to herding models. Like herding models, our model assumes that choices between alternatives are determined by the opinion of others. The crucial difference, however, is that our model leaves out one of the key component of information cascade models: the assumption that people rely on popularity to make inferences about the quality of alternatives.

To clarify the difference between our sampling model and herding and information cascade models, suppose that a majority of individuals come to believe that A is superior to B. Herding models assume that information about such majority opinions directly impacts quality estimates: people infer that A is likely superior to B. Because information about the majority opinion influence quality estimates (people infer A is superior) it also has an indirect effect on choice behavior (people choose A because they believe it is superior). Our model does not assume that information about the majority opinion directly influences quality estimates. In our model, the majority opinion influences behavior (people likely sample A). Quality estimates are only indirectly influenced by majority opinions through its effect on sampling (because people sample A their negative opinions about A could change). Stated differently, in our model popularity only has an indirect effect on beliefs; an effect that is mediated by the effect on popularity on sampling.

Because our model does not assume that information about the majority opinion directly influences quality estimates it may be more suitable for settings in which such an assumption is not warranted. First, individuals may have different tastes or skills and may not believe that others
know best (e.g. Banerjee et al., 2013; Hendricks et al., 2012; Munshi, 2004). In situations where tastes and skills differ between people, it makes little sense for an individual to make an inference about how good an alternative is for her on the basis of popularity. For example, a vegetarian should probably not pay much attention to the popularity of restaurants among non-vegetarians. Although it does not make sense, in these situations, for the decision maker to infer how good an alternative will be for her on the basis of popularity, our model can still apply. To see how, note that even when tastes or skills differ, people might still have a tendency to select popular alternatives. Vegetarians might be motivated to go to a restaurant liked by non-vegetarians to spend time with friends. In such settings, the payoff distributions of the alternatives systematically differ across people but they still have a tendency to select popular alternatives. To see how we can analyze such situations, consider a variant of our model with binary payoffs where the only modification is that we now suppose that the success probabilities of the alternatives are specific to each individual. Let $p^i_k$ be the success probability of Alternative $k$ for Individual $i$. The proofs of our main results (Proposition 1 and 2) can be easily adapted to this case (at the cost of some complications in the notations). In other words, our theory predicts that quality estimates will tend to synchronize even when there is taste and skill heterogeneity. This is in contrast to the predictions of information cascade accounts of conformity. Since it does not make sense to use popularity to infer quality in these situations, these theories would not predict that beliefs would become synchronized.

Second, our theory also fits settings where information about more popular alternatives is more immediately available and nudges people into selecting them without people forming conscious impressions about the quality of different alternatives. This frequently happens on online recommendation systems embedded in review websites (e.g. TripAdvisor, Yelp, Zagat) or media streaming platforms (e.g. Youtube, Spotify) where items liked by many are often displayed at the top of search results. Readers who browse the web by clicking on hyperlinks are thus likely to sample again opinions that are already popular in their (online) social circles (Sunstein, 2009). Our model suggests that such subtle influences on sampling behavior have the potential to create a synchronization of opinions, even if people do not make inferences about the worthiness of opinions on the basis of their popularities. In fact, our model does not require that people be aware of the fact that some alternatives are more popular than others. All that matters is that they have an increased propensity to sample again alternatives liked by others.
6 Application to Social Choice

The crucial assumption of our model is that people are more likely to sample popular alternatives than unpopular alternatives. An important reason for why this often happens is social choice, such as majority voting. When groups select alternatives by using majority voting (Davis 1973; Hastie & Kameda, 2005; Stasser, Kerr & Davis, 1980), every member will be exposed to the more popular alternative, even those who were in favor of another alternative. Our theory predicts that such social choice rule will contribute to an homogenization of opinions.

To see how, consider members of a university department who vote about new appointments on a periodic basis. When a new faculty member joins, all the members of the department get exposed to the research tradition the new member represents. Based on this experience, members can update their attitudes toward this research tradition. Our theory implies that the outcome of this process of learning and voting is an homogenization of attitudes: most people will come to prefer the same research tradition. Although this prediction would not be surprising in settings where all the group members made the same observations, our theory also predicts that voting will lead to an homogenization of attitudes even if the observations of the group members are independent from each others, i.e. are independent realizations of the underlying random variable. This implies, in particular, that attitudes will tend to become more homogenous even if group members have different tastes.

To demonstrate this formally, consider a variant of our model with binary payoffs. All the assumptions remain unchanged except for the sampling rule: we now assume that, in every period, all the individuals sample the most popular alternative. If \( N_1 > N_2 \) \( (N_1 < N_2) \), everybody selects Alt. 1 (Alt. 2). If \( N_1 = N_2 \), one alternative is selected randomly and everybody samples it. The observations of all the individuals are independent realizations of the underlying payoff variable. With these assumptions, the stationary distribution of the popularities is in fact very similar to that of the model with binary payoffs (Proposition 1).\(^{10}\) Thus, quality estimates will become

\(^{10}\)The omitted proof is almost identical to the proof of Proposition 1. The joint stationary distribution of the sum is similar to eq. 6, by taking \( s = \infty \):

\[
\tilde{f}(n_1,n_2) = K \left( I_{n_1=0} + I_{n_2=0} \right) (p_1)^{n_1} (1 - p_1)^{n-n_1} (p_2)^{n_2} (1 - p_2)^{n-n_2} \binom{n}{n_1} \binom{n}{n_2},
\]

where \( I_{n_j=0} \) is an indicator function equal to 1 when \( n_j = 0 \) and equal to 0, otherwise.
synchronized: most individuals will agree that one alternative is superior to the other one, but the favored alternative does not have to be the superior one.

In summary, majority voting combined with individual learning can lead to an homogenization of opinions, possibly in favor of a less-than-optimal alternative. Unlike prior literature, our model does not need to assume that group members identify with the group (e.g. Mackie, 1986). As such, our theory casts new light on the conditions where group membership contributes to attitude and opinion homogeneity.

7 Discussion and Conclusion

Our model offers an existence proof that a sampling based account could explain phenomena usually attributed to either rational popularity-based inferences or to irrational cognitive biases. Our approach does not require that people engage in sophisticated inferences based on the choices of others or that people engage in motivated cognition. Rather, it emphasizes basic learning processes in an environment where sampling is driven by the opinions of the majority. Sampling influenced by the majority offers a simple alternative explanation for how public conformity in behavior can lead to private acceptance and a possible collective illusion.

How can the sampling-based account proposed in this paper be empirically distinguished from the inference-based mechanism? To address this question, it is useful to consider a setting where fine-grained data about quality estimates are available and popularity affects both quality inferences and sampling behavior. Suppose that at the beginning of each period, people report their quality estimates for the two alternatives A and B ($\hat{q}_{t}^{A}$ and $\hat{q}_{t}^{B}$), then learn about the popularity of the alternatives ($N_{t}^{A}$ and $N_{t}^{B}$), then report their quality estimates given what they know about the popularities (let us call these $\hat{q}_{t}^{A\ast}$ and $\hat{q}_{t}^{B\ast}$ – these quantities are not part of our formal model), then sample one of the alternatives, then report their quality estimates again ($\hat{q}_{t+1}^{A}$ and $\hat{q}_{t+1}^{B}$). Consider alternative A. In this context, the effect of popularity-based inferences is the effect of popularity on $\hat{q}_{t\ast}^{A} - \hat{q}_{t}^{A}$. The effect of popularity through sampling is the effect of popularity on $\hat{q}_{t+1}^{A} - \hat{q}_{t\ast}^{A}$. The total effect of popularity on quality estimates is the effect of popularity on $\hat{q}_{t+1}^{A} - \hat{q}_{t}^{A}$. If there is no popularity-based quality inferences, we have $\hat{q}_{t\ast}^{A} = \hat{q}_{t}^{A}$. In this case, the effect of popularity would be entirely driven by its effect on sampling behavior, as in our formal model. In the remaining of
This discussion we refer to this case as the ‘direct sampling account’.

This example shows that the ‘direct sampling account’ has empirical implications distinct from inference-based accounts. To cleanly distinguish between these accounts, however, researchers need data about quality estimates after people learn about the popularity and before they make their sampling decisions (i.e. researchers need $\hat{q}_{t}^{A}$ and $\hat{q}_{t}^{B}$). In many naturally occurring environments such detailed data on quality estimates is not available. Researchers may have access to data on quality estimates and sampling behavior in each of several periods, but may not know if quality estimates do reflect, or not, the information agents have about the popularities of the alternatives. With such data, researchers cannot distinguish between the direct effect of popularity on sampling behavior and its indirect effect via quality estimates. That is, suppose a researcher finds that popularity in period $t - 1$ affects sampling behavior in period $t$. Such an effect may occur because popularity directly influences sampling (as in the ‘direct sampling account’) or because popularity indirectly influence sampling via popularity-based quality inferences.

Our sampling model can be useful even when it is not possible to empirically disentangle the two ways in which popularity has an effect on sampling behavior. The sampling model shows that the effect of popularity on sampling behavior - whether mediated by quality estimates or not - must be taken into account to explain the association between popularity and quality estimates. Researchers who ignore the effect of popularity on sampling will overestimate the effect of popularity on quality estimates through popularity-based inferences. In addition to any direct effect of popularity on estimates - due to inferences about quality from popularity – our model shows that in settings where people learn about the qualities of the alternatives partly from their own experiences, there is also an indirect effect of popularity on beliefs that operates via sampling behavior.

This paper has focused on collective illusions but a sampling account could be applied to other phenomena in information economics where prior theoretical accounts have emphasized inferences from choices. Consider signaling (Spence, 1973). Models of costly signaling assume that audiences make inferences about individual traits on the basis on choices of activities. A sampling account suggests that these choices can also impact preferences over activities: people choosing education sample information about this activity and learn that they enjoy education. More generally, economic incentives that influence choice will likely have systematic effects on evaluations and beliefs because - as the empirical literature on sampling effects shows - sampling biases have systematic
effects on belief formation (Fiedler, 2000; Fazio, Eiser, and Shook, 2004; Fetchenhauer and Dunning, 2010).

References


A Appendix 1: Proofs for the Model with Binary Payoffs

A.1 Proof of Proposition 1

$N_1$ and $N_2$ are the random variables $N_{1,t}$ and $N_{2,t}$ converge to. The state space is of size $(2 \times 2)^n$.
The Markov process, $\{N_{1,t}, N_{2,t}, t = 1, 2, \ldots\}$, is irreducible and aperiodic. Therefore, the joint distribution of $(N_1, N_2)$ is uniquely defined (Lemma 5.14 in Ross & Pekoz, 2007).

We first derive the stationary joint distribution of estimates of all the individuals in the model.
Then we use this distribution to derive the stationary distribution of the popularities. Let \( \{ \hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}_n, \hat{Q}_2 \} \) be the random variables toward which the quality estimates of the \( n \) individuals in the population \( (\{ \hat{Q}_{1,t}, \hat{Q}_{2,t}, \ldots, \hat{Q}_n^{1,t}, \hat{Q}_2^{1,t} \}) \) converge.

**Lemma 1.** The stationary joint density of the estimates of all the individuals (the \( \hat{q}_1 \)'s and \( \hat{q}_2 \)'s can be equal to one or zero) is equal to:

\[
f(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2) = K(e^{-sn_1} + e^{-sn_2})p_1^{n_1}(1-p_1)^{n-n_1}p_2^{n_2}(1-p_2)^{n-n_2}
\]  

(6)

where \( n_1 = \sum_{j=1}^{n} \hat{q}_i^j \) and \( n_2 = \sum_{j=1}^{n} \hat{q}_j^j \) and \( K^{-1} = (1-p_1 + p_1e^{-s})^n + (1-p_2 + p_2e^{-s})^n \).

**Proof.** The Markov chain has a unique stationary distribution. To prove that the distribution defined in eq. 6 is the stationary distribution, it is thus enough to prove that it satisfies the following stability equation:\(^{11}\)

\[
f(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2) = \sum_{k=1}^{4} B_k(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2),
\]  

(7)

with

\[
B_1(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2) = \sum_{j=1}^{n} \frac{1}{n} f(q_1, q_2, \ldots, 1, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2)pC1p_1^{q_1}(1-p_1)^{1-q_1},
\]  

(8)

\[
B_2(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2) = \sum_{j=1}^{n} \frac{1}{n} f(q_1, q_2, \ldots, 1, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2)pC1p_1^{q_1}(1-p_1)^{1-q_1},
\]  

(9)

\[
B_3(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2) = \sum_{j=1}^{n} \frac{1}{n} f(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2)(1-pC1)p_2^{q_2}(1-p_2)^{1-q_2},
\]  

(10)

\[
B_4(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2) = \sum_{j=1}^{n} \frac{1}{n} f(q_1, q_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2)(1-pC1)p_2^{q_2}(1-p_2)^{1-q_2},
\]  

(11)

To explain this, consider the summand of \( B_1 \). This is the probability that individual \( j \) gets selected \((1/n)\), that the quality estimates equal \( \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2 \) (which, by hypothesis, occurs with probability \( f(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_1, \hat{q}_2) \)), that Individual \( j \) samples the Alt. 1 (occurs

\(^{11}\)This is assuming that \( 0 < n_1, n_2 < n \). What follows can easily be adapted to the cases where \( n_1 \) or \( n_2 \) is equal to 0 or \( n \).
with probability $pC_1$), and that her estimate does not change. When $\hat{q}^j_i = 1$, her estimate does not change when $j$ gets a success (this occurs with probability $p_1$); when $\hat{q}^j_i = 0$, her estimate does not change when $j$ gets a failure (this occurs with probability $1 - p_1$).

The summand of $B_2$ is the probability that individual $j$ gets selected, that the quality estimates equal $\hat{q}_1^j, \hat{q}_2^j, ..., 1 - \hat{q}_1^j, \hat{q}_2^j, ..., \hat{q}_1^n, \hat{q}_2^n$, that she samples Alt. 1 (this occurs with probability $pC_1$), and that her estimate changes. When $1 - \hat{q}^j_i = 1$, a change happens when $j$ gets a failure (this occurs with probability $1 - p_1 = (1 - p_1)^{1 - \hat{q}^j_i}$); when $1 - \hat{q}^j_i = 0$, a change happens when $j$ gets a success (this occurs with probability $p_1 = (p_1)^{\hat{q}^j_i}$).

Similar explanations apply to the third and fourth sums at the right hand side of the stability equation.

Now, we verify that the formula for the joint distribution in eq. 6 satisfies the above stability equation.

Using the expression for $f(\hat{q}_1^j, \hat{q}_2^j, ..., \hat{q}_1^n, \hat{q}_2^n)$ given by eq. 6, and noting that $pC_1(n_1, n_2) = 1/(1 + e^{s(n_1-n_2)}) = e^{-sn_2}/(e^{-sn_2} + e^{-sn_1})$, the term which is summed up in $B_1$ can be expressed as

$$K e^{-sn_2} p_1^{n_1+\hat{q}^j_i} (1 - p_1)^{n-n_1+1-\hat{q}^j_i} p_2^{n_2} (1 - p_2)^{n-n_2}.$$ 

Similarly, the term which is summed up in $B_2$ can be expressed as

$$K e^{-sn_2} p_1^{n_1+1-\hat{q}^j_i} (1 - p_1)^{n-n_1+\hat{q}^j_i} p_2^{n_2} (1 - p_2)^{n-n_2}.$$ 

Then we have

$$B_1 + B_2 = K \sum_{j=1}^n \frac{1}{n} e^{-sn_2} p_1^{n_1} (1 - p_1)^{n-n_1} p_2^{n_2} (1 - p_2)^{n-n_2} \left[ p_1^{\hat{q}^j_i} (1 - p_1)^{1-\hat{q}^j_i} + p_1^{1-\hat{q}^j_i} (1 - p_1)^{\hat{q}^j_i} \right].$$

However, the term within the bracket always equals 1, for if $\hat{q}^j_i = 1$ then

$$p_1^{\hat{q}^j_i} (1 - p_1)^{1-\hat{q}^j_i} + p_1^{1-\hat{q}^j_i} (1 - p_1)^{\hat{q}^j_i} = p_1 + 1 - p_1 = 1. \quad (12)$$

36
And if \( \hat{q}_i^j = 0 \), then
\[
p_1^{\hat{q}_i^j}(1 - p_1)^{1-\hat{q}_i^j} + p_1^{1-\hat{q}_i^j}(1 - p_1)^{\hat{q}_i^j} = 1 - p_1 + p_1 = 1.
\] (13)

So,
\[
B_1 + B_2 = K \sum_{j=1}^{n} \frac{1}{n} e^{-sn_2} p_1^{n_1} (1 - p_1)^{n-n_1} p_2^{n_2} (1 - p_2)^{n-n_2}.
\] (14)

A similar reasoning gives
\[
B_3 + B_4 = K \sum_{j=1}^{n} \frac{1}{n} e^{-sn_1} p_1^{n_1} (1 - p_1)^{n-n_1} p_2^{n_2} (1 - p_2)^{n-n_2}.
\] (15)

Finally,
\[
\sum_{k=1}^{4} B_k = K \sum_{j=1}^{n} \frac{1}{n} (e^{-sn_1} + e^{-sn_2}) p_1^{n_1} (1 - p_1)^{n-n_1} p_2^{n_2} (1 - p_2)^{n-n_2}
\]
\[
= K (e^{-sn_1} + e^{-sn_2}) p_1^{n_1} (1 - p_1)^{n-n_1} p_2^{n_2} (1 - p_2)^{n-n_2}.
\] (16)

which is the expression for the joint estimate density in eq. 6. Also, it is easy to verify that with the proposed formula for \( K \), the expression in eq. 6 sums to 1. This concludes the proof.

Given the joint distributions of estimates, it is easy to get the joint distribution of the sums, \( \bar{f}(n_1, n_2) \). First, what is the probability of one way of assigning beliefs so that the sums are \( n_1 \) and \( n_2 \)? From equation 6 we know that it is given by
\[
f(\hat{q}_1^1, \hat{q}_2^1, ..., \hat{q}_1^n, \hat{q}_2^n) = K (e^{-sn_1} + e^{-sn_2}) p_1^{n_1} (1 - p_1)^{n-n_1} p_2^{n_2} (1 - p_2)^{n-n_2}.
\] (17)

Second, the number of ways there are to have \( \sum_{j=1}^{n} \hat{q}_1^j = n_1 \) and simultaneously \( \sum_{j=1}^{n} \hat{q}_2^j = n_2 \) is \( \binom{n}{n_1} \binom{n}{n_2} \). By multiplying, we get \( f(n_1, n_2) \).

A.2 Proof of Proposition 2

(i) Suppose, without loss of generality, that \( k = 1, i = 1 \) and \( j = 2 \). In order to show that \( \text{corr} (\hat{Q}_1^1, \hat{Q}_1^2) > 0 \), it is enough to show that the joint distribution of estimates, \( f \), is totally
positive of order 2 with respect to its first and third arguments \((TP_2, \text{see Karlin and Rinott, 1980, p. 470-471, eq. 1.7-1.9})\). \(f\) is \(TP_2\) with respect to its first and third arguments if and only if, for all \(\hat{q}_1^1, \hat{q}_1^{1*}, \hat{q}_2^1, \hat{q}_1^2, \hat{q}_2^2, \hat{q}_1^3, \hat{q}_2^3, ..., \hat{q}_1^n, \hat{q}_2^n\) such that \(\hat{q}_1^1 < \hat{q}_1^{1*}\) and \(\hat{q}_1^2 < \hat{q}_2^{2*}\), we have

\[
f(\hat{q}_1^1, \hat{q}_2, \hat{q}_1^{2*}, \hat{q}_2^2, \hat{q}_1^3, \hat{q}_2^3, ..., \hat{q}_1^n, \hat{q}_2^n)f(\hat{q}_1^1, \hat{q}_2, \hat{q}_1^2, \hat{q}_2^2, \hat{q}_1^3, \hat{q}_2^3, ..., \hat{q}_1^n, \hat{q}_2^n) > f(\hat{q}_1^{1*}, \hat{q}_2^1, \hat{q}_1^{2*}, \hat{q}_2^2, \hat{q}_1^3, \hat{q}_2^3, ..., \hat{q}_1^n, \hat{q}_2^n) \quad (19)
\]

Let \(\hat{q}_1^1, \hat{q}_1^{1*}, \hat{q}_1^2, \hat{q}_1^{2*}\) be such that \(\hat{q}_1^1 < \hat{q}_1^{1*}\) and \(\hat{q}_1^2 < \hat{q}_2^{2*}\). Given the restrictions on the quality estimates in this model, we have \(\hat{q}_1^1 = 0, \hat{q}_1^{1*} = 1, \hat{q}_1^2 = 0\) and \(\hat{q}_1^{2*} = 1\). To simplify the formulas, let denote \(\sum_{j=3}^n \hat{q}_j^1\) by \(n_1^{3-n}\). Applied to the asymptotic joint density, equation 18 becomes:

\[
K \left( e^{-s(1+1+n_1^{3-n})} + e^{-sn_2} \right) p_{1+1+n_1^{3-n}}^{n-(1+1+n_1^{3-n})} p_{2}^{n_2} (1-p_2)^{n-n_2} \\
K \left( e^{-s(0+0+n_1^{3-n})} + e^{-sn_2} \right) p_{0+0+n_1^{3-n}}^{n-(0+0+n_1^{3-n})} p_{2}^{n_2} (1-p_2)^{n-n_2} \\
> K \left( e^{-s(1+0+n_1^{3-n})} + e^{-sn_2} \right) p_{1+0+n_1^{3-n}}^{1+(1+0+n_1^{3-n})} (1-p_1)^{n-(1+0+n_1^{3-n})} p_{2}^{n_2} (1-p_2)^{n-n_2} \\
K \left( e^{-s(0+1+n_1^{3-n})} + e^{-sn_2} \right) p_{0+1+n_1^{3-n}}^{0+(1+0+n_1^{3-n})} (1-p_1)^{n-(0+1+n_1^{3-n})} p_{2}^{n_2} (1-p_2)^{n-n_2}
\]

Simple algebraic manipulations show that this inequality is equivalent to:

\[
\left( e^{-s(2+n_1^{3-n})} + e^{-sn_2} \right) \left( e^{-sn_1^{3-n}} + e^{-sn_2} \right) > \left( e^{-s(1+n_1^{3-n})} + e^{-sn_2} \right) \left( e^{-s(1+n_1^{3-n})} + e^{-sn_2} \right) \quad (19)
\]

Further manipulations show that this inequality holds true iff

\[
e^{-2s} + e^{-0s} - 2e^{-s} > 0 \quad (20)
\]

The LHS of this equation can be rewritten as \((1 - e^{-s})^2\). This quantity is positive because \(s > 0\). This implies that the inequality in eq. 19 is true. In conclusion, \(f\) is \(TP_2\) with respect to its first and third arguments. This implies that \(corr \left( \hat{Q}_1^1, \hat{Q}_1^2 \right) > 0\).

\((ii)\) This proposition follows by a reasoning similar to that used for the proof of \((i)\).
A.3 Proof of Corollary 1

Without loss of generality, consider $k = 1$, $i = 1$, and $j = 2$. Because the density, $f$, has a multiplicative form, eq. 18 implies

$$\frac{P(\hat{Q}_1^1 = 1, \hat{Q}_2^2 = 1)}{P(\hat{Q}_1^1 = 0, \hat{Q}_2^2 = 1)} > \frac{P(\hat{Q}_1^1 = 1, \hat{Q}_2^2 = 0)}{P(\hat{Q}_1^1 = 0, \hat{Q}_2^2 = 0)}. \quad (21)$$

By dividing the numerator and denominator of the LHS by $P(\hat{Q}_2^2 = 1)$, we obtain the LHS of the inequality of the statement of the corollary. By dividing the numerator and denominator of the RHS, by $P(\hat{Q}_2^2 = 0)$, we obtain the RHS of the statement of the corollary.

B Proofs for the Model with Normally Distributed Payoffs

We will use the following notations: $\hat{Q}_1 = (\hat{Q}_1^1, ..., \hat{Q}_1^n)$, $\hat{Q}_2 = (\hat{Q}_2^1, ..., \hat{Q}_2^n)$, $\hat{Z}_i = (\hat{Q}_i^1, \hat{Q}_i^2)$ and $\hat{Z} = (\hat{Z}_1, ..., \hat{Z}_n)$. $\hat{Z}$ is the random variable $\hat{Z}_t$ converges to. We first compute the joint distribution of estimates. It is formalized in the following lemma:

**Lemma 2.** The joint stationary distribution of the quality estimates of all individuals in the population for the two alternatives is

$$h(\hat{Z}) = K \left( e^{-\frac{s^2}{b} \sum_{i=1}^n \hat{q}_i^1} + e^{-\frac{s^2}{b} \sum_{i=1}^n \hat{q}_i^2} \prod_{i=1}^n g_1(\hat{q}_i^1) g_2(\hat{q}_i^2) \right), \quad (22)$$

where

$$K^{-1} = e^{\frac{s^2 \mu_2}{b(2-\mu)}} (e^{-s\mu_1} + e^{-s\mu_2}) \quad (23)$$

is a normalizing constant, and $g_j(\cdot)$ is the density of the estimate that an individual would converge to if she always sampled Alt. $j$.

In this model, the ‘popularity’ of an alternative is the sum of the quality estimates of that alternative in the population. Let $\overline{Q}_1 = \sum_i \hat{Q}_i^1$ and $\overline{Q}_2 = \sum_i \hat{Q}_i^2$ and let $h(\cdot, \cdot)$ denote the joint distribution of $\overline{Q}_1$ and $\overline{Q}_2$. The joint distribution of the popularities is given by the following lemma:

**Lemma 3.** [Joint Stationary Distribution of Sums of Estimates] The stationary distribution of the...
sufficient to the quality estimates, $Q_1$ and $Q_2$ is:

$$h(q_1, q_2) = K(e^{-s\frac{q_1}{n}} + e^{-s\frac{q_2}{n}})h_1(q_1)h_2(q_2)$$  \tag{24}$$

where, for $j \in \{1, 2\}$, $h_j(\cdot)$ is the density of normal distribution with mean $n\mu_j$ and variance $n\sigma^2 b/(2 - b)$ and $K$ is a normalizing constant.

The joint distribution is a weighted bivariate Normal distribution where the weight is $K(e^{-s\frac{q_1}{n}} + e^{-s\frac{q_2}{n}})$. As in the above model, the weighting function puts additional weight on the extremes where the alternatives are unpopular. And thus the joint distribution of popularities is bimodal if $s$ is high enough. This implies that a collective illusion might emerge in which the suboptimal alternative is evaluated more positively than the superior alternative.

### B.1 Proof of Lemma 2

We first need to show that the stationary distribution of estimates exists. Because the model defines a Markov process in general state space (non-countable), this is a rather involved proof. Fortunately, we can use some recent developments on the analysis of the stability of Harris chains to get to our objective.

#### B.1.1 Convergence of the Markov Process

With the above specification, the (random) vector of estimates, $\hat{Z}_t$, defines a Harris chain that is positive recurrent. This implies that $\hat{Z}_t$ has a unique stationary distribution $h$ (e.g. Meyn & Tweedie, 1993). To prove this and the convergence, we use the following theorem (Glynn, 2011).

**Theorem.** Consider a Markov chain $R = (R_t : t \geq 1)$ with state space $S$. Suppose there exists $A \subseteq S$, $\gamma > 0$, $m \geq 1$ and a distribution $\varphi$ on $S$, $\epsilon > 0$ and a function $g : S \to [0, \infty]$ such that:

1. $E_r g(R_2) \leq g(r) - \epsilon$, $r \in A^C$;
2. $\sup_{r \in A} E_r g(R_2) < \infty$;
3. $P_r(R_m \in \cdot) \geq \gamma \varphi(\cdot)$, $r \in A$.

Then $R$ possesses a unique stationary distribution $h$ and for each non-negative $f$,
\[ \frac{1}{n} \sum_{k=1}^{n} f(R_t) \to \int f(r)h(dr) \text{ a.s.} \]

We now verify that the Markov process \((\hat{Z}_t : t \geq 1)\) satisfies the conditions of the theorem. In our setting, the state space is \(\mathbb{R}^{2n}\).

i) Let

\[ g(\hat{z}_t) = \sum_{i=1}^{n} (|\hat{q}_{1,i}^i| - \mu_1 + |\hat{q}_{2,i}^i| - \mu_2) \]

and \(A = ([\mu_1 - 1, \mu_1 + 1] \times [\mu_2 - 1, \mu_2 + 1])^n\). Let \(\hat{z}_1 \in A^c\). We have

\[ E_{\hat{z}_1} g(\hat{Z}_2) = \sum_{i=1}^{n} \frac{1}{n} pC_1(\hat{z}_1) \left[ (1 - b) |\hat{q}_{1,1}^i - \mu_1| + bE |a_{1,1}^i - \mu_1| + \sum_{j \neq i} |\hat{q}_{1,j}^i - \mu_1| + \sum_{j} |\hat{q}_{2,j}^i - \mu_2| \right] \]

\[ + \sum_{i=1}^{n} \frac{1}{n} pC_2(\hat{z}_1) \left[ \sum_j |\hat{q}_{1,j}^i - \mu_1| + (1 - b) |\hat{q}_{2,j}^i - \mu_2| + bE |a_{2,j}^i - \mu_2| + \sum_{j \neq i} |\hat{q}_{2,j}^i - \mu_2| \right] \quad (25) \]

Noting that for all \(i\), \(E |a_{1,1}^i - \mu_1| = 0\) and \(E |a_{2,j}^i - \mu_2| = 0\). Then

\[ E_{\hat{z}_1} g(\hat{Z}_2) = \sum_{i=1}^{n} \frac{1}{n} \left[ \sum_j |\hat{q}_{1,j}^i - \mu_1| + \sum_j |\hat{q}_{2,j}^i - \mu_2| \right] \]

\[ + \frac{b}{n} \sum_{i=1}^{n} \left[ pC_1(\hat{z}_1) |\hat{q}_{1,1}^i - \mu_1| + pC_2(\hat{z}_1) |\hat{q}_{2,1}^i - \mu_2| \right], \quad (26) \]

\[ = \sum_{i=1}^{n} \left[ |\hat{q}_{1,1}^i - \mu_1| + |\hat{q}_{2,1}^i - \mu_2| \right] \]

\[ - \frac{b}{n} \sum_{i=1}^{n} \left[ pC_1(\hat{z}_1) |\hat{q}_{1,1}^i - \mu_1| + pC_2(\hat{z}_1) |\hat{q}_{2,1}^i - \mu_2| \right], \quad (27) \]

\[ = g(\hat{z}_1) - \frac{b}{n} \sum_{i=1}^{n} \left[ pC_1(\hat{z}_1) |\hat{q}_{1,1}^i - \mu_1| + pC_2(\hat{z}_1) |\hat{q}_{2,1}^i - \mu_2| \right]. \quad (28) \]

The fact that \(\hat{z}_1 \in A^c\) and the definition of \(A\) imply that for all \(i\), \(|\hat{q}_{1,1}^i - \mu_1| > 1\) and \(|\hat{q}_{2,1}^i - \mu_2| > 1\). Therefore,

\[ \sum_{i=1}^{n} \left[ pC_1(\hat{z}_1) |\hat{q}_{1,1}^i - \mu_1| + pC_2(\hat{z}_1) |\hat{q}_{2,1}^i - \mu_2| \right] > \sum_{i=1}^{n} \left[ pC_1(\hat{z}_1) + pC_2(\hat{z}_1) \right] = n. \quad (29) \]

Finally, \(E_{\hat{z}_1} g(\hat{Z}_2) < g(\hat{z}_1) - b\). Setting \(\epsilon = b > 0\) ensures that condition i is satisfied.
\[ \text{Let } \hat{z}_1 \in A. \text{ We have} \]

\[ E_{\hat{z}_1} g(\hat{Z}_2) = g(\hat{z}_1) - \frac{\lambda}{N} \sum_{i=1}^{N} \left[ pC_1(\hat{z}_1) \left| \hat{q}^i_{1,1} - \mu_1 \right| + pC_2(\hat{z}_1) \left| \hat{q}^i_{2,1} - \mu_2 \right| \right] \leq g(\hat{z}_1). \]  

(30)

Then, \( \sup_{\hat{z}_1 \in A} E_{\hat{z}_1} g(\hat{Z}_2) \leq \sup_{\hat{z}_1 \in A} g(\hat{z}_1) < \infty \), the last inequality being a consequence of the fact that \( A \) is a bounded space.

\[ \text{iii) Let } B \subseteq \mathbb{R}^{2n}. \text{ In } 2n \text{ transitions, the process can reach any point in the state space } \mathbb{R}^{2n} \text{ from any starting point. This is the result of a combination of two facts: First, there is a positive probability that all individuals will have been selected and will have each chosen once the first alternative and the second alternative and thus there is a positive probability that all estimates will have been updated. Second, because updated estimates are weighted averages of previous estimates and normally distributed observations, the support of the joint distribution of updated estimates is the whole state space } \mathbb{R}^{2n}. \text{ This implies that if we write} \]

\[ P_{\hat{z}_1}(\hat{Z}_{2n+1} \in B) = \int_B p_{2n}(\hat{z}_1, r) dr, \]  

(31)

then the transition density \( p_{2n}(\cdot, r) \) is positive. It is also continuous, as a weighted sum of normal densities.

We can then define \( \phi(r) = \inf_{\hat{z}_1 \in A} p_{2n}(\hat{z}_1, r) \). The facts that \( p_{2n} : \mathbb{R}^{2n} \times \mathbb{R}^{2n} \rightarrow \mathbb{R} \) is continuous and that \( A \) is a compact set imply that \( \phi \) is continuous. Let \( \varphi(B) = \int_B \phi(r) dr \), and set \( \gamma = 1 \).

With these definitions, for all \( \hat{z}_1 \in A, P_{\hat{z}_1}(\hat{Z}_{2n+1} \in \cdot) \geq \gamma \varphi(\cdot) \), which proves that condition \( iii \) is satisfied.

In conclusion, the above theorem implies that there exists a unique stationary joint distribution for the estimates.

\[ \text{B.1.2 The Stationary Distribution of the Quality Estimates} \]

The density \( g_j(\cdot) \) satisfies the following equality

\[ \int_r g_j(r)p_j(r, x) dr = g_j(x), \]  

(32)
where \( p_j(r, x) \) is the probability density of the event that the estimate of individual \( i \) transitions from \( r \) in period \( t \) to \( x \) in period \( t + 1 \) given that individual \( i \) samples Alt. \( j \) in period \( t + 1 \). Under the above assumptions, \( p_j(r, x) \) equals \( \frac{1}{b} f_j \left( \frac{1}{b} x - (1 - b)r \right) \), where \( f_j(\cdot) \) is the normal density of the observation of Alt. \( j \) and \( g_j(\cdot) \) is a normal density with mean \( \mu_j \) and variance \( \sigma^2 b/(2 - b) \).

The joint probability density of estimates has to satisfy the following stability equation

\[
\frac{1}{n} \sum_{k=1}^{n} \frac{1}{n} \int_{r_k} h(q_1^1, q_2^1, ..., r_k, q_2^k, ..., q_1^n, q_2^n) pC_1(q_1^1, q_2^1, ..., r_k, q_2^k, ..., q_1^n, q_2^n) p_1(r_k, q_1^k) dr_k
+ \sum_{k=1}^{n} \frac{1}{n} \int_{r_k} h(q_1^1, q_2^1, ..., q_1^k, r_k, ..., q_1^n, q_2^n) pC_2(q_1^1, q_2^1, ..., q_1^k, r_k, ..., q_1^n, q_2^n) p_2(r_k, q_2^k) dr_k. \tag{33}
\]

To explain this, consider the first term in the first sum:

\[
\frac{1}{n} \int_{r_k} h(q_1^1, q_2^1, ..., r_k, q_2^k, ..., q_1^n, q_2^n) pC_1(q_1^1, q_2^1, ..., r_k, q_2^k, ..., q_1^n, q_2^n) p_1(r_k, q_1^k) dr_k.
\]

This is the probability that individual \( k \) gets selected, that all other individuals have estimates \( \hat{q}_1^1, \hat{q}_2^1, ..., \hat{q}_1^n, \hat{q}_2^n \), that individual \( k \) has estimates equal to \( r_k \) and \( \hat{q}_2^k \), that she samples the first alternative, and that her estimate goes from \( r_k \) to \( \hat{q}_1^k \) which happens with probability \( p_1(r_k, \hat{q}_1^k) \).

We integrate over all possible estimates about the first alternative. The other terms have a similar explanation. To show that the above joint density (eq. 22) does satisfy the stability equation (eq. 33) note first that \( pC_1(z) \) can be written as follows

\[
pC_1(\hat{z}) = \frac{e^{-s \frac{1}{n} \sum_{i=1}^{n} \hat{q}_2^i}}{e^{-s \frac{1}{n} \sum_{i=1}^{n} \hat{q}_1^i + e^{-s \frac{1}{n} \sum_{i=1}^{n} \hat{q}_2^i}}}. \tag{34}
\]

We insert this expression and the stationary joint density into the first term of the first sum:

\[
\frac{1}{n} \int_{r_k} h(q_1^1, q_2^1, ..., r_k, q_2^k, ..., q_1^n, q_2^n) pC_1(q_1^1, q_2^1, ..., r_k, q_2^k, ..., q_1^n, q_2^n) p_1(r_k, q_1^k) dr_k
\]

\[
= K \frac{1}{n} e^{-s \frac{1}{n} \sum \hat{q}_2^i} \prod_{i \neq k} g_1(\hat{q}_1^i) \prod_{i} g_2(\hat{q}_2^i) \int_{r_k} g_1(r_k) p_1(r_k, \hat{q}_1^k) dr_k
\]

\[
= K \frac{1}{n} e^{-s \frac{1}{n} \sum \hat{q}_2^i} \prod_{i \neq k} g_1(\hat{q}_1^i) \prod_{i} g_2(\hat{q}_2^i). \tag{35}
\]

The last equality is a consequence of the fact that, by definition of \( g_1, \int_{r_i} g_1(r_i) p_1(r_i, \hat{q}_1^i) dr_i = g_1(\hat{q}_1^i) \).
Similarly, the first term in the second sum is equal to:

$$K \frac{1}{n} e^{-s \frac{1}{n} \sum_j \hat{q}^1_j} \prod_j g_1(\hat{q}^1_j) \prod_j g_2(\hat{q}^2_j).$$

(36)

Summing across all terms, the right hand side of equation 33 becomes equal to

$$K \left( e^{-s \frac{1}{n} \sum_i \hat{q}^1_i} + e^{-s \frac{1}{n} \sum_i \hat{q}^2_i} \right) \prod_i g_1(\hat{q}^1_i) g_2(\hat{q}^2_i),$$

(37)

which is \( h(\hat{q}^1_1, \hat{q}^1_2, ..., \hat{q}^n_1, \hat{q}^n_2) \), the stationary joint density.

Noting that the integral of the density should be equal to 1, we obtain the formula for \( K \). We use it to write an explicit formula for \( h \):

$$h(\hat{z}) = \left( \frac{1}{\sqrt{2\pi \frac{b}{b-n} \sigma^2}} \right)^{2n} e^{-s \frac{1}{n} \sum_i \hat{q}^1_i + e^{-s \frac{1}{n} \sum_i \hat{q}^2_i}} e^{-\sum_i \frac{(\hat{q}^1_i - \mu_1)^2 + (\hat{q}^2_i - \mu_2)^2}{2 \sigma^2}}.$$  

(38)

**B.2 Proof of Lemma 3: Joint Distribution of the Sum of Estimates**

To get the joint distribution of the sum, we need to perform a complicated integration of the joint density (eq. 22) over the region in which the sums equal some constants such as \( \sum_{i=1}^n \hat{Q}^1_i = \overline{q_1} \) and \( \sum_{i=1}^n \hat{Q}^2_i = \overline{q_2} \). However, note that if the sums equal \( \overline{q_1} \) and \( \overline{q_2} \), the first part in the density always equals \( (e^{-s \overline{q_1} / n} + e^{-s \overline{q_2} / n}) \). Thus, we can ignore the first part.

The second part is then simple - what remains is a sum of independent normal distributions! The distribution of the sum of the estimates is thus a normal distribution with mean equal to the sum of the means and variance equal to the sum of the variances.

**B.3 Proof of Proposition 3**

(i) The formula follows by calculating \( E(\hat{Q}^1_k) \), \( E\left( (\hat{Q}^1_k)^2 \right) \) and \( E(\hat{Q}^1_k \hat{Q}^2_k) \) by doing the appropriate integrations on the basis of eq. 22. The calculations are tedious but straightforward, we therefore only provide only the key steps. We have:

$$E(\hat{Q}^1_k) = \mu_k - \frac{s \sigma^2 b}{n(2-b)} \frac{e^{-s \mu_k}}{e^{-s \mu_1} + e^{-s \mu_2}}.$$
The proof follows by noting that
\[
E \left( \hat{Q}^i_k \hat{Q}^j_k \right) = (\mu_k)^2 - \frac{s^2 b}{n(2-b)} \frac{e^{-s\mu_k}}{e^{-s\mu_1} + e^{-s\mu_2}} \left( 2\mu_k - \frac{s^2 b}{n(2-b)} \right)
\]
\[
E \left( \left( \hat{Q}^i_k \right)^2 \right) = (\mu_k)^2 + \frac{s^2 b}{2-b} + \frac{e^{-s\mu_k}}{e^{-s\mu_1} + e^{-s\mu_2}} \left( \frac{s^2 b}{n(2-b)} \right)^2 - 2 \frac{s^2 b}{n(2-b)} \mu_k
\]

The proof follows by noting that \( \text{cov} \left( \hat{Q}^i_k, \hat{Q}^j_k \right) = E \left( \hat{Q}^i_k \hat{Q}^j_k \right) - E \left( \hat{Q}^i_k \right) E \left( \hat{Q}^j_k \right) \), \( \text{var} \left( \hat{Q}^i_k \right) = E \left( \left( \hat{Q}^i_k \right)^2 \right) - \left( E \left( \hat{Q}^i_k \right) \right)^2 \), \( \text{corr} \left( \hat{Q}^i_k, \hat{Q}^j_k \right) = \text{cov} \left( \hat{Q}^i_k, \hat{Q}^j_k \right) / \sqrt{\text{var} \left( \hat{Q}^i_k \right) \text{var} \left( \hat{Q}^j_k \right)} \) and performing the appropriate substitutions and simplifications.

(ii) We have:
\[
E \left( \hat{Q}^i_1 \hat{Q}^j_2 \right) = \mu_1 \mu_2 - \frac{s^2 b}{n(2-b)} \frac{\mu_2 e^{-s\mu_1} + \mu_1 e^{-s\mu_2}}{e^{-s\mu_1} + e^{-s\mu_2}}
\]

The proof follows from computations similar to those in the proof of (i).