CASCADE EFFECTS IN HETEROGENEOUS POPULATIONS

Daniel B. Neill

ABSTRACT

We present a model of sequential choice which explains the emergence and persistence of unpopular, inefficient behavioral norms in society. We model individuals as *naïve Bayesian norm followers*, rational agents whose subjective expected utility is increased by adherence to an established norm. Agents use Bayesian reasoning to combine their private preferences and prior beliefs with empirical observations of others' decisions. When agents must infer the preferences of others from observation, this can result in *negative cascades*, causing the majority of agents to choose a dispreferred action (because they believe, incorrectly, that they are following the majority preference). We demonstrate that negative cascades can result even when the degree of conformity is relatively low, and under a wide range of conditions (including heterogeneity in preferences, priors, and impact of public opinion). This allows us to present a general model of how rational norm-following behavior can occur, and how unpopular norms might emerge, in real populations with heterogeneous preferences and beliefs.

KEY WORDS: • cascades • herding • naïve Bayes • norm followers

1. Introduction

It is well known that the opinions of others play a huge role in many of the everyday decisions we make. The clothes we wear, the cars we drive, and the food we eat are all influenced to varying extents by public opinion; even our major life decisions such as which schools to attend, or what career to pursue, are strongly dependent on the choices of others. The role of 'peer pressure' in influencing the behavior of schoolchildren is well established: children may be motivated either to take drugs, or to 'just say no,' depending on the influence of their peers. Nor are adults immune to this pressure: our support of a

Rationality and Society Copyright © 2005 Sage Publications. Vol. 17(2): 191–241.
www.sagepublications.com DOI: 10.1177/1043463105051633

political candidate, or our stance on a public issue, may be influenced by opinion polls or by the views of our neighbors. The 'bandwagon' effect (Leibenstein 1950) is well known in the marketing literature: consumers are much more likely to buy a product if they are convinced of its popularity. This is why, for example, McDonald's advertises 'billions and billions' of hamburgers sold, and why companies such as Nike and Reebok spend millions of dollars annually on celebrity endorsements.

The phenomenon of 'herd behavior' occurs when people follow the actions of others, even when their private information or preferences suggest an alternative course of action. Beginning with the seminal work of Banerjee (1992), Bikhchandani et al. (1992), and others, various models have been proposed to explain how herding can result from individual rational decisions. These models have been applied to a variety of political, social, and economic situations, including product choices, investment, voting, fertility choices, political movements, fashions, fads, and cultural change. In particular, in sequential choice situations where individuals are influenced by the actions of prior decision-makers, various positive feedback mechanisms can trigger *cascade effects* where it is optimal for each new decision-maker to follow the behavior of the preceding individuals, regardless of his own private information or preferences.[1]

A number of positive feedback mechanisms have been identified and discussed in the literature; in many cases, multiple mechanisms may contribute to a cascade effect. In the work of Katz and Shapiro (1986) on technology choice, positive feedback results from *network externalities*: widespread adoption of a single technology reduces the costs due to lack of compatibility between different technologies. This can lead to lock-in of an early-established technology, making it difficult for later alternatives to gain market share even if the alternatives are technologically superior; one example of this is the QWERTY keyboard design (Dawid 1985). Brian Arthur's (1989) work on increasing returns and lock-in in technology choice discusses these effects as well as *price and performance externalities*. Increased investment in a given technology generally leads to advances in that technology, resulting in lower prices and improved performance; also, increased production decreases unit cost due to economies of scale, and these mechanisms encourage further investment in the technology at the expense of alternative technologies. Arthur's examples of lock-in due to price and performance externalities include the adoption of the gasoline-powered automobile (rather than steam-powered alternatives) in the 1890s; another example would be the triumph of the microcomputer against minicomputer and mainframe alternatives. More generally, *coordination externalities* can result in many decision situations where an individual's payoff is affected by the actions of the others with whom he interacts; these include 'collective action' such as political demonstrations (Lohmann 1994) as well as social dilemmas such as the Prisoner's Dilemma.

Perhaps the most well known of these positive feedback mechanisms is found on the literature on 'informational cascades' (Bikhchandani et al. 1992; Banerjee 1992). The informational cascade model describes situations where individuals must choose between several alternatives, one of which is 'objectively correct' (and thus, would be preferred by all individuals if they had complete information). However, each individual only has partial information (a noisy signal which may be positively correlated with the correct choice), and thus each must choose based on his private signal and his observation of the choices of others. In sequential choice situations, informational cascades arise when it is optimal for an individual (having observed the actions of prior decision-makers) to follow the behavior of the preceding individuals without regard to his own information. As a result, the decisions of individuals in a cascade convey no new information, and thus each new decision-maker (presented with an identical decision situation) also chooses to ignore his own information and follow the herd. Thus one of the most important results in the herding literature is a demonstration of the power of informational cascades in inducing conformity: in models such as Bikhchandani et al. and Banerjee, informational cascades will always arise, and unless the cascade is broken by external shocks (such as the release of public information), it will continue to induce conforming behavior indefinitely.

1.1 Informational vs. Reputational Cascades

Informational cascades have been used to explain a wide variety of phenomena, including investment in financial markets (e.g. Welch 1992; Lee 1998), fashions, fads, and cultural change. For some of these phenomena, the informational cascade model is clearly appropriate: for example, the laboratory experiments of Anderson and Holt (1997), where each individual in turn must guess (based on partial information) from which urn a ball has been drawn. Similarly, in the investment literature, the question is what action a firm should take to maximize its profit: thus one of the choices is objectively correct, and players must use the information provided by other players' choices to maximize their probability of making the correct choice. However, in other examples of herding behavior (such as product choice), the 'correct' choice is *subjective*, and may differ from person to person. In our canonical example (a choice between two brands of soft drink), each person's optimal choice depends both on his own private preference as well as his desire to choose the more popular brand.

Thus it is clear that, in these subjective choice situations, it is not informational cascades but other positive feedback mechanisms which cause cascade effects.[2] Many of these feedback mechanisms can be grouped under the heading of *reputational effects*: effects resulting from the approval or disapproval of others. As discussed in depth by Kuran (1995), reputational concerns can have huge effects on decision-making, and can often cause individuals to make decisions contrary to their private preferences, a phenomenon known as *preference falsification*. For example, Scharfstein and Stein (1990) show that reputational effects can cause investment managers to follow the herd, even when they believe that they could have made greater expected profit by acting differently. Some reputational effects may be direct: societies or individuals may punish deviations from an established norm, with the severity of sanctions ranging from mild disapproval (negative expressions or words) to severe punishment (including torture or death). Similarly, adherence to a norm may lead to approval or even substantial rewards. In many cases, norms and customs have been sufficiently internalized that deviating from the norm may cause severe emotional stress even in the absence of external sanctions. Even in cases where there is no clearly established societal norm, and neither rewards nor punishments are likely to be substantial (such as our soft drink example), individuals still tend to follow the perceived majority preference. One reason for this is the strong psychological connection between reputation and perceived self-worth: individuals often have a strong fear of social criticism or rejection, and even mild disapproval by a peer group may cause severe damage to self-esteem.

As Kuran (1995) notes, however, it is often impossible to please everyone. Preferences that are acceptable to one individual or group may be unacceptable to another, and thus decision-makers must balance the reputational effects resulting from approval/

disapproval from many different sources. Thus Kuran defines reputational utility as the net payoff generated by the sum of positive and negative sanctions for holding that preference. As a result, reputational effects may be dependent not only on the majority preference but on the size of that majority: an individual may choose to act in opposition to a slight majority, while choosing to follow the norm set by a larger majority. In cases where groups have differing amounts of power and prestige, this discrepancy must also be taken into account; a dominant social group may have a significant impact on societal norms even if representing only a small share of the population.

1.2 Strong vs. Weak Cascades

In the literature, cascade effects are typically labeled by the primary mechanism of positive feedback: thus we have ‘informational cascades’ resulting from informational effects, ‘reputational cascades’ resulting from reputational effects, and similarly ‘networking,’ ‘price,’ or ‘availability’ cascades. For our formal model of cascade effects, however, our main interest is not the positive feedback mechanism but the effect it creates: how agents combine their private preference with their perception of the distribution of preferences in society, in order to make a decision. We distinguish between two types of cascades, which we denote by *strong* and *weak cascades* respectively. A strong cascade occurs when an agent’s private preference is weighted proportional to the preference of other individuals, and a weak cascade occurs when an agent’s private preference is weighted proportional to the preference of society as a whole. The two types of cascade tend to exhibit very different long-run behavior: in strong cascades, once a large number of other individuals’ decisions have been observed, an agent’s private preference is almost certain to be ‘drowned out’ by the information conveyed by previous decisions. Thus *total cascades* typically occur: there exists some point after which every individual will follow the herd, and thus (in the limit of an infinite population) the proportion of herd-followers tends to 1. In weak cascades, on the other hand, the number of prior decision-makers does not matter, but only the perceived distribution of preferences in the population. Thus individuals make different choices depending on the strength of their private preferences, as compared to the size of the majority. As noted above, some individuals may choose to follow a norm if they believe that it is followed by nearly everyone in a society, but may diverge from the norm if a significant minority chooses to do so as well. As a result, *partial cascades* typically occur: the long-run proportion of individuals following the norm will converge to some value $p < 1$ as the population size grows large. In other words, in a weak cascade, every individual *whose reliance on public opinion is sufficiently high* will follow the herd, but other individuals may choose to diverge from the herd; the long-run result of the cascade depends on the distribution of weights of public opinion in the population. It is clear from this description that strong cascades tend to be easier to analyze than weak cascades, and the majority of the literature has focused on strong cascades. Our model focuses on the weak cascade case, but can also be easily applied to strong cascades; as a result, many of the strong cascade models (including Arthur 1989; Bikhchandani et al. 1992; and Bicchieri and Fukui 1999) can be represented as special cases of our model. This is discussed in more detail in Section 4.4.

The clearest example of a strong cascade is the typical informational cascade setting of Banerjee (1992) and Bikhchandani et al. (1992). In this setting, any individual will follow the herd once the total information conveyed by previous individuals' choices outweighs his own private information, and thus total cascades occur. If every individual has identical precision of information, an individual's private information would be weighted proportional to the information of a single observed decision; if individuals have greater confidence in their own information than that of others, private information may be given a higher weight. A second example of strong cascades is the reputational cascade setting of Bicchieri and Fukui (1999). In this case, all individuals are highly sensitive to public opinion: individuals are assumed to follow the perceived majority preference (if one exists) regardless of the size of the majority. Thus an individual's private preference is counted as one 'vote' in determining the majority preference; this model can be easily extended to allow individuals to place extra weight on their private preferences. In this case, the reputational cascade setting becomes formally identical to an informational cascade model, with weight of private preference corresponding to precision of private information: in fact, we can view this as an informational cascade where the goal is to infer, and follow, the majority preference. A third and final example of strong cascades can be found in Arthur's (1989) work on price and performance externalities, as discussed above. Economies of scale, and improvements due to technological advance, are typically influenced by the total production of a technology rather than its relative proportion in the population. Thus, if the technology is made sufficiently reliable and inexpensive, almost everyone will prefer this technology over alternatives, and total cascades will occur. This is clear from Arthur's model, where after lock-in of a product occurs, it is in everyone's best interest to choose the locked-in technology.

The clearest example of a weak cascade is the case of a reputational cascade with no clear societal norm, and no strong or explicit sanctions, as in our canonical example of soft drink choice. In this case, each individual or group may have a different preference, and a decision-maker must consider the distribution of preferences across the population. In such cases, a slight majority may not be sufficient to cause an individual to choose contrary to his private preference, but a large majority may sway the individual to follow the herd. More generally, weak cascades may arise from a variety of coordination effects, where utility is increased by actions or preferences that 'agree' with those of others, and decreased by actions or preferences that 'disagree' with those of others. Then total utility from coordination effects is computed as a sum of the utilities for agreement and disagreement with another individuals, weighted by the probabilities of agreement/disagreement. This is similar to, but more general than, the reputational utility of Kuran (1995): coordination effects can result in utility due to gain or loss of reputation, but can also affect utility in a variety of other ways. One example is the networking cascades of Katz and Shapiro (1986). In this case, the payoffs do not result from reputation, but from compatibility; choosing identical technologies allows easier communication and interaction between the possessors of these technologies, while choosing different and incompatible technologies may make mutually beneficial interactions difficult or impossible to achieve.

1.3 Inefficiency of Cascade Effects

Another important result of the literature on cascade effects, and a major focus of this paper, is that cascades (though caused by individually rational behavior) may lead to inefficient outcomes. Banerjee (1992) presents a model of informational cascades with a continuum of possible choices and a single correct choice; he proves that, for any size of the population, the probability that no one in the population chooses the correct option is bounded away from zero. Bikhchandani et al. (1992) and Welch (1992) consider binary choices, and show that informational cascades can cause behavior to converge to the wrong choice. Bikhchandani et al. comment that 'even for very informative signals . . . the probability of the wrong cascade is remarkably high.' The reason for this inefficiency is that cascades prevent aggregation of information: once a cascade begins, individuals' actions convey no information about their private signals or preferences, and thus do not improve the quality of later decisions. Thus, if early decisions convey incorrect or misleading information, the cascade effects prevent (or at least, make more difficult) the dissemination of new and correct information, amplifying the impact of the incorrect information on future decisions. Similarly, in the work of Katz and Shapiro (1986) and Arthur (1989) on technology choice, random events (amplified by positive feedback mechanisms such as network and price externalities) can cause an inferior technology to be 'locked-in,' preventing the dissemination of superior alternatives. As Arthur notes, we should be very cautious of any explanation that seeks to explain adoption of a technology in terms of the winner's 'innate superiority,' since it is often the inferior technology which is adopted.

Bicchieri and Fukui (1999) have used cascade effects to explain the emergence and persistence of unpopular and inefficient norms in society. Examples of such norms include widespread corruption and bribery of public officials (Bicchieri and Rovelli 1995), violent behavior by members of juvenile gangs, and norms of discrimination against minorities (Bicchieri and Fukui 1999); in each case, the norms persist despite being dispreferred by the majority of society, and even by the groups actively involved in maintaining the norm. Several factors could contribute to the persistence of unpopular norms: individuals may choose irrationally, or in social dilemmas such as the Prisoner's Dilemma, individually rational choices may result in poor collective outcomes. However, Bicchieri and Fukui have shown that unpopular norms can result in a population of rational agents, even in interactions where there is no conflict between the individual and collective good. This can occur whenever individual choices are influenced by the preferences of others, but the exact distribution of preferences is not known and must be inferred from observation. If the behavior that an individual observes does not reflect the true preferences of society as a whole, he may be influenced to choose an action which is dispreferred both by himself and by society. As a result, others will observe his action and adjust their choices accordingly: this can lead to a *negative cascade* in which the majority of people choose a dispreferred action. For example, in our canonical example of soft drink preference, an individual may drink Pepsi instead of Coke because he observes others drinking Pepsi and assumes that Pepsi is universally preferred, though in fact those others are only drinking Pepsi because they observed others doing so, and so on. In this paper, as in much of the prior literature on cascade effects, one major goal is to examine the probability that negative cascades will occur, and the magnitude of the effect of these cascades, in a variety of sequential decision-making situations. Our 'biased random walk' model allows exact computation of these probabilities and magnitudes under certain simplifying assumptions, and precise estimation of these quantities by simulation in the more general model setting.

1.4 Our Model

Thus we present a model of cascade effects which builds on the prior work of Bicchieri and Fukui (1999), as well as much of the earlier work on informational and reputational cascades. As in other models of cascade effects, we examine a *micromodel* of the behavior of individual rational agents, and show that this leads to herd behavior in the aggregate. Moreover, we show that herding is often inefficient, in that (with finite and often high probability) the majority of the population will choose an action which is contrary to their private preference. As in the cascade models of Arthur (1989), Banerjee (1992), and Bikhchandani et al. (1992), we typically assume that the order of moves is sequential, exogenous, and randomly determined. However, we also examine situations where there is some simultaneity in decision-making; this is discussed in Section 4.6.

Our goal is much the same as that of Bicchieri and Fukui (1999): to examine the prevalence of inefficient and unpopular norms in society through quantitative investigation of cascade effects in general, and negative cascades in particular. However, we reject some of the simplifying assumptions made by their model (and much of the prior literature) in order to present a more general and realistic model of cascade effects in heterogeneous populations. First, it should be noted that our model is not limited to one particular positive feedback mechanism (such as 'informational cascades' or 'reputational cascades') but can be applied across all of these models. To achieve cascade effects, only two conditions are necessary: sequential choice with observation of previously made choices, and some sort of positive feedback mechanism which encourages individuals to follow the herd. Then the essential question is how individuals *combine* their observations of previously made choices with their private preferences and prior beliefs. We propose a two-level hierarchical model: first, individuals use 'naïve Bayesian' reasoning to combine their observations and priors into an estimate of the distribution of preferences across the population. Second, individuals combine this estimate with their private preference using an additive utility model similar to (but distinct from) the dual preference model of Kuran (1995). As is evident from this method, our focus is on 'weak cascades' (since individuals weight their private preference proportional to the distribution of preferences in society as a whole), but the model can also be easily adapted to 'strong cascades,' as discussed in Section 4.4. Most importantly, in the weak cascade setting, we do not require every member of the population to follow the perceived majority preference, but allow them to take into account the size of the majority as compared to the strength of their private preferences. This is very different than previous cascade models such as Banerjee (1992), Bikhchandani et al. (1992), and Bicchieri and Fukui (1999), which assume that every individual places very high weight on public opinion, and thus follows the majority. In such situations, it is not surprising that cascades occur, but the more realistic scenario (where many individuals do not place high weight on public opinion) has not been adequately explored. In this scenario, our model demonstrates that (in the limit of a large population) 'partial cascades' will always occur. Moreover, in many cases these cascades will be inefficient: negative cascades can occur even in a population where the majority of individuals put low weight on popular opinion. However, the impact of a cascade may differ depending on individuals' weights of public opinion: in some cases, cascades may have small or negligible effects, while in other cases they may exert a huge influence on decisions.

This brings us to the most interesting aspect of our model of cascade effects: the possibility of a heterogeneous population. In addition to having different private preferences, individuals may have different prior beliefs about the distribution of preferences in the population, and may be influenced to differing extents by their private preferences, prior beliefs, and observations of other people's behavior. As a result, each individual may have a different 'threshold' for following the herd, where the threshold may be based both on the number of prior decision makers and the observed proportions of each choice. This concept of threshold is similar to that of Granovetter and Soong (Granovetter 1978; Granovetter and Soong 1983, 1986, 1988). Our model differs from Granovetter's, however, in both the micromodel (how the thresholds are computed) and the macromodel (aggregation of individual choices into cascade effects). In particular, in our model each agent's threshold results from combining private preferences, priors, and observations in a certain way (given by our micromodel of 'naïve Bayesian norm followers') while Granovetter and Soong simply assume a distribution of thresholds across the population and examine the resulting dynamics. Additionally, our macromodel of sequential decisions is stochastic (modeled as a biased random walk), while Granovetter and Soong use deterministic models including a simple bandwagon effect (each individual acts once his threshold is passed) and population modeling by differential and difference equations. Thus the main difference of our model from the prior literature is its applicability to heterogeneous populations, where individuals differ not only in their preferences and prior beliefs, but also in the relative weights they give to their private preferences, priors, and observations. An individual in our model is characterized by four continuous parameters: his norm-independent utility, weight of popular opinion, weight of observation, and prior; many of the previous models are special cases of this, where all individuals have high weights of public opinion, high weights of observation, and equal priors.

Thus we present a sequential choice model of decision-making, and show that cascade effects emerge from individual rational decisions. We model individual decision-makers as 'naïve Bayesian norm followers,' rational but myopic agents who use Bayesian reasoning to combine their private preferences and prior beliefs with their empirical observations of other agents' decisions. A naïve Bayesian norm follower calculates the expected utility of an action as a weighted sum of its 'norm-independent utility' (NIU) and its 'norm-following utility' (NFU). The NIU of an action is independent of the choices of other agents, while the NFU of an action is proportional to the agent's Bayesian estimate of the probability that the action will agree with that of another randomly selected agent. In cases where disagreement between agents has potential negative consequences, it may be in an individual's best interest to choose actions which he believes are favored by the majority, even if the actions are dispreferred with respect to his private preferences. But in cases where a) there is no significant majority, b) there has not been enough observed behavior to determine a majority, or c) the influence of public opinion is low, the individual may choose his preferred action even if this may cause disagreement with others.

In the following sections, we present our model of cascade effects. Section 2 presents the micromodel, examining how individual naïve Bayesian norm followers make preference decisions. Section 3 examines the aggregate behavior of a population of naïve Bayesian norm followers, using a 'biased random walk' model to calculate the probability and magnitude of (positive and negative) cascades. Section 4 generalizes the model to preference-dependent parameters, unequal priors, and varying weights of observation and public opinion. Section 5 discusses the underlying assumptions of myopic 'naïve Bayesian' rationality in the micromodel, and Section 6 concludes the paper.

2. A Model of Naïve Bayesian Norm Followers

Our micromodel of individual rational decision-making attempts to answer the question of how an individual combines three distinct influences: his private preference (i.e. which action he would prefer if the preferences of other members of the population were irrelevant), his prior beliefs (i.e. his prior expectation of the distribution of preferences in the population), and his observations of others' choices. We propose a two-level hierarchical model. At the top level, individuals are 'norm followers': rational decision-makers who maximize subjective expected utility, where the utility of an action is assumed to be increased both a) if it reflects the agent's private preference, and b) if it adheres to an established norm (or majority preference). At the lower level, individuals use 'naïve Bayesian' reasoning to combine their observations and priors into an estimate of the distribution of preferences across the population. Combining these two levels appropriately, we obtain our model of 'naïve Bayesian norm followers.'

At the top level of our model, the total utility of an action is the sum of two components: a norm-independent utility (NIU) and a norm-following utility (NFU). The norm-independent utility is a measure of the agent's private preference for an action: it can be thought of as a reward or punishment which the agent receives regardless of the actions of other members of the population. For example, if an agent prefers Pepsi to Coke (in the absence of information about others' preferences), this is equivalent to saying that NIU(Pepsi) is higher than NIU(Coke) for that agent. The norm-following utility of an action is independent of the agent's private preference: instead, it is a measure of the agent's desire to choose actions which are preferred by the majority of others (i.e. to follow a norm if one is established). For example, NFU(Pepsi) would be higher than NFU(Coke) for an agent if the agent believes that the majority of people prefer Pepsi, and if this belief makes the agent more likely to choose Pepsi. This is similar to the dual preference model of Kuran (1995), where an individual's public preference is computed by choosing the option which maximizes the sum of 'intrinsic,' 'reputational,' and 'expressive' utilities. Kuran's model is specifically geared toward the case of a public vote without secret balloting: the 'intrinsic utility' of an action reflects its impact on the decision of society as a whole, the 'reputational utility' of an action represents the influence of others through reputational effects, and the 'expressive utility' of an action reflects an individual's desire to express his private preference (avoiding preference falsification). Our model is geared more toward individual, subjective preference decisions, rather than votes for the preference of society as a whole. For example, an individual does not drink Coke because we wants to transform the world into a Coke-drinking society; he drinks Coke because he enjoys it. Thus our norm-independent utility is somewhat similar to Kuran's expressive utility: an individual maximizes norm-independent utility by choosing the action corresponding to his private preference (though simply because he prefers this option, not because he feels the general need to express autonomy). Similarly, our norm-following utility includes not only reputational effects, but all the effects of other individuals' choices, including informational and coordination effects. The important issues in our model are that the individual desires both to follow his private preference and to follow the norm: thus we distinguish between these 'norm-independent' and 'norm-following' components of total utility.

Why might a rational agent prefer to follow a norm? In the previous section, we discussed a variety of positive feedback mechanisms (including information, reputation, and networking/coordination) which influence individuals to follow the observed behavior of the herd. For another useful perspective on norm-following, we turn to Bicchieri (2003), who presents a detailed discussion of the different types of behavioral norms and why a rational agent might choose to follow each. She distinguishes between 'descriptive norms,' which are the equilibria of coordination games, and 'social norms,' which transform social dilemmas (e.g. the Prisoner's Dilemma) into coordination games. In brief, an individual would wish to follow a descriptive norm because his utility is maximized by coordinating with others: for example, he would prefer to drive on the right side of the road if he believed that others did the same. For a behavioral rule R to be a descriptive norm, a rational agent will prefer R conditional on his belief that others prefer R. For social norms, on the other hand, an agent's preference for R is conditional not only on his belief that others prefer R, but also on his belief that others expect him to prefer R. Thus an agent might wish to follow a social norm for a number of reasons: he might seek approval (or reward, or agreement) for following the norm, he might seek to avoid disapproval (or sanctioning, or disagreement) for failure to follow the norm, or he might internalize the norm (and as a result, his private preferences would shift toward the norm). These reasons for norm-following fit in well with our discussion of positive feedback mechanisms above, including both coordination and reputational effects.

For the purposes of our model, we define a norm as a perceived regularity in behavior that, in turn, influences behavior via some positive feedback mechanism. Norms need not be universally followed to have an influence on behavior: rather, we assume that agents seek to maximize the *probability* of agreement with others. This desire is rational since we are assuming that some positive feedback mechanism (whether information, reputation, or coordination) is at work, and thus an agent's utility is increased by agreement with the perceived majority preference. The norm-following utility of an action a is defined to be a weighted average of the utilities for agreeing and disagreeing with other agents, where each utility is weighted by its probability of occurrence. Let us assume a binary decision between two actions, A and B. Then an agent's norm-following utility for choosing action A is $\mathrm{NFU}(A) = u_+P(A) + u_-P(B)$, where u_+ and u_- are the utilities for agreement and disagreement respectively ($u_+ > u_-$), and $P(A)$ and $P(B)$ are the agent's estimates of the probabilities that another agent will choose A or B respectively. From this, we can calculate his total subjective expected utility for choosing A: $U(A) = \text{NIU}(A) + \text{NFU}(A) = u_A + u_+P(A) + u_-P(B)$. Similarly, the agent's total utility for choosing B is: $U(B) = \text{NIU}(B) + \text{NFU}(B) = u_B + u_+P(B) + u_-P(A)$. Then a rational agent will choose A whenever $U(A) > U(B)$, that is, whenever $U(A) - U(B) = (u_A - u_B) + (P(A) - P(B))(u_+ - u_-) > 0$.

Now we must consider the second level of our hierarchical model, which explains how the agent estimates the probabilities $P(A)$ and $P(B)$. A 'naïve Bayesian norm follower' will estimate these probabilities by using Bayes' rule to combine his observations with his prior estimates of these probabilities. To do so, he assumes that choices are binomially distributed, and thus infers the *maximum a-posteriori estimate* of $P(A)$ using a beta-binomial model. Assuming that the agent has observed N actions, N_A of which were action A, and N_B of which were action B, and assuming that the agent has a Beta(α, β) prior,[3] he estimates:

$$P(A) = \arg\max_\theta \Pr(\theta|N_A, N_B) = \arg\max_\theta \Pr(N_A, N_B|\theta)\Pr(\theta)$$

$$= \arg\max_\theta \theta^{N_A}(1-\theta)^{N_B}\theta^{\alpha-1}(1-\theta)^{\beta-1} = \frac{N_A + (\alpha - 1)}{N + (\alpha + \beta - 2)}$$

A more useful parametrization of this expression is in terms of $P_0(A)$, the agent's *prior estimate* of P(A), and k, the agent's *weight of observation*. Setting $P_0(A) = (\alpha - 1)/(\alpha + \beta - 2)$ and $k = (1)/(\alpha + \beta - 2)$, we obtain $P(A) = (P_0(A) + kN_A)/(1 + kN)$.[4] Thus, under the beta-binomial model, the Bayesian posterior estimate of $P(A)$ simplifies to a weighted sum of the prior estimate $P_0(A)$ and the empirical (observed) proportion N_A/N. Similarly, we can compute $P(B) = (P_0(B) + kN_B)/(1 + kN)$. Note that $P(A) + P(B) = 1$, so the Bayesian posterior estimates are consistent. The weight of observation k is a positive constant which is assumed to be fixed for a given agent, but can vary from agent to agent: agents with a high weight of observation are more influenced by their empirical observations (i.e. the proportion of observed actions in which A is chosen), while agents with a low weight of observation are more influenced by their prior estimate of the probability of A.

Now we can combine the two levels of our hierarchical model. Using the agent's Bayesian estimates for $P(A)$ and $P(B)$, in the expression derived for $U(A) - U(B)$, we obtain:

$$U(A) - U(B) = (u_A - u_B) + \frac{P_0(A) - P_0(B) + k(N_A - N_B)}{1 + kN}(u_+ - u_-)$$

Now we let $\Delta u = u_A - u_B$, $\Delta P = P_0(A) - P_0(B)$, $\Delta N = N_A - N_B$, and $q = u_+ - u_-$. This gives us:

$$U(A) - U(B) = \Delta u + q\frac{\Delta P + k\Delta N}{1 + kN}$$

A rational agent will choose action A when $U(A) - U(B) > 0$, choose action B when $U(A) - U(B) < 0$, and be indifferent between A and B when $U(A) - U(B) = 0$. Thus the agent's choice is dependent on six variables. Two of the variables represent the current state: the total number of actions observed N, and the difference between the numbers of A and B actions observed, ΔN. The other four variables are parameters intrinsic to a given agent: his private preference (parametrized by Δu), his prior (parametrized by ΔP), his weight of observation k, and his weight of public opinion q. We assume that each agent's utility function is normalized so that $|\Delta u| = 1$; we can do this without loss of generality since an agent's utility function is only unique up to a positive linear transformation (von Neumann and Morgenstern 1944), and no inter-agent utility comparisons are made. We also note that the quantity $\left|(\Delta P + k\Delta N)/(1 + kN)\right| \leq 1$. Thus, if the (normalized) weight of public opinion $q \gg 1$, the individual relies heavily on public opinion to make his choices, and if $q < 1$, the individual is not influenced by the preferences of others. For moderate values of q, the individual will only be influenced by public opinion if the proportion of the majority is sufficiently large.

Now we must ask an essential question: what makes our naïve Bayesian norm followers 'naïve'? This is not meant to be a pejorative term, but rather a descriptive one: they follow a decision rule similar to that of a 'naïve Bayes classifier' in the machine learning literature (e.g. Mitchell 1997), maximizing Bayesian a-posteriori probability under the assumption that all observations are independent given the model. It is important to note, however, that this assumption of independence of observations is blatantly false: in fact, the sequence of observations exhibits strong dependence due to cascade effects, and the naïve Bayesian norm followers do not take this into account. We revisit this issue in detail in Section 5, considering a variety of reasons why agents might fail to consider cascades, and thus exhibit naïveté.[5]

3. Naïve Bayesian Norm Followers and Cascade Effects

We now consider a large population of naïve Bayesian norm followers, each of whom must choose between the two actions A and B. We assume without loss of generality that action A is preferred by the majority of the population, if a majority exists: let $x \geq .5$ be the proportion of the population preferring A. Each agent must estimate x based on his observations of other agents' actions: agents do not know the size of the majority, nor do they know which action is preferred by the majority (i.e. they do not know that $x \geq .5$). As derived above, an agent will choose A when $\Delta u + q(\Delta P + k\Delta N)/(1 + kN) > 0$, where Δu, ΔP, q, and k are parameters intrinsic to that individual. Thus we can ask the very general question: what proportion of agents will choose A, given the distribution of these four parameters in the population? This depends, of course, on the information available to each player at the time when he makes his decision. If all players decide simultaneously, an agent will choose A if $\Delta u + q\Delta P > 0$, and thus the proportion of the population choosing A would simply be the proportion for which this relation holds. For instance, if all agents have $\Delta P = 0$ (equal prior probabilities for A and B), then each agent would choose according to his private preference, and thus the proportion of the population choosing A would be equal to x. Much more interesting effects occur when the players choose sequentially and can observe the choices of previous decision-makers; we focus on this *sequential choice model* for the remainder of the paper. We typically assume that the order of decisions is randomly determined, that players can observe all previous decisions, and choose one at a time.[6] Assuming a sufficiently large population size, this is equivalent to a model where a randomly selected player is chosen from the population each turn, with values of Δu, ΔP, q and k drawn at random from their respective distributions. This player will then choose A or B based on his four parameters as well as the current values of ΔN and N. As a result of this choice, N will be increased by one, and ΔN will be increased or decreased by one (if the player chooses A or B respectively). Thus we can model the sequential choice situation as a *random walk* on ΔN, where N is an additional state variable which influences the transition probabilities. Unfortunately, no general closed form solution exists for this random walk model, for arbitrary distributions of Δu, ΔP, q, and k. However, we can easily simulate such a random walk, maintaining the current values of ΔN and N: on each time step, we sample the four parameters from their respective distributions, and adjust ΔN and N accordingly. In Sections 4.5 and 4.6 we consider simulation results for the general model; for the remainder of the paper, however, we make various simplifying assumptions which allow calculation of a closed form solution. We begin by considering an extremely simple model, and gradually remove these simplifying assumptions through this section and the next, enabling us to examine the behavior of our model under a wide variety of conditions, including heterogeneity in preferences, priors, and the weights of observation and public opinion.

We initially consider the simple case where all agents have $|\Delta u| = 1$ (normalized payoffs), $\Delta P = 0$ (equal prior probabilities for A and B) and $k \gg 1$ (high weight of observation). This allows us to simplify the decision rule: an agent will choose action A when $\Delta u + q(\Delta N)/(N + \varepsilon) > 0$, where $\varepsilon = 1/k$ is a small positive constant. Thus an agent who prefers action A will make his dispreferred choice if $N \geq 1$ and $\Delta N < -N/q$. Similarly, an agent who prefers action B will make his dispreferred choice if $N \geq 1$ and $\Delta N > N/q$. For now, we make the further simplifying assumption that each member of the population is either strongly dependent on public opinion ($q \gg 1$) or makes decisions independently of public opinion ($q < 1$). In this case, an individual with $q \gg 1$ chooses A when $\Delta N > 0$, chooses B when $\Delta N < 0$, and chooses his private preference when $\Delta N = 0$. An individual with $q < 1$ chooses his private preference regardless of ΔN. Let z equal the proportion of individuals who are strongly dependent on public opinion; we initially assume that an individual's dependence on public opinion is independent of his preference for either A or B. Now we ask the following question: in terms of x (the proportion of the population preferring A) and z (the proportion of the population dependent on public opinion), what proportion of the population will actually choose action A?

Though in our general model, each player observes all previous actions, let us first consider an even simpler model, where each player can observe only the previous player's action. In this case, we can easily calculate a closed-form solution: we find that the prob-

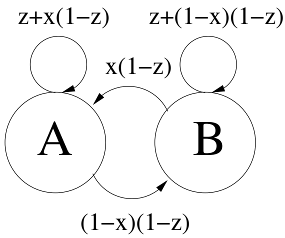

Figure 1. Markov chain for sequential choice dependent on previous player's action

ability of a randomly selected player choosing A is $z + x(1 - z)$ if the previous player chose A, and $x(1 - z)$ if the previous player chose B. Thus we have a two-state Markov chain with transition probabilities shown in Figure 1, and we can calculate the stationary distribution of this chain. For any $z < 1$, we find that the proportion of the population choosing A converges to x. In other words, the same proportion of the population chooses A as if all players had chosen simultaneously, so no cascade effect has occurred. However, if $z = 1$, then every player will make the same choice as the first player: thus with probability x every player will choose A, and with probability $1 - x$ every player will choose B. In either of these situations, a *total cascade* has occurred: every individual, in turn, chooses to follow the observed behavior of others regardless of his own private preference. The first situation, where everyone in the population chooses A, is known as a *positive cascade*; the second situation, where everyone in the population chooses B, is known as a *negative cascade*. More generally, a positive cascade occurs when the proportion of the population choosing the majority-preferred action is greater than the size of the majority; a negative cascade occurs when the proportion of the population choosing the majority-preferred action is less than the size of the majority. We are most interested in negative cascades which cause the majority of the population to choose a dispreferred action, resulting in an inefficient and unpopular behavioral norm.

3.1 The Biased Random Walk Model

It is not surprising, of course, that cascade effects can occur when every individual's behavior is determined by public opinion (i.e.

when $z = 1$). With a slight adjustment to our model, we find a much more interesting variety of results. Thus we consider the sequential choice model discussed above, where each player can observe the aggregate result of all previous choices: the jth player to make a choice knows how many of the first $j - 1$ players chose A and B respectively, and can use this information to estimate the probability $P(A)$ with Bayes' rule.

To calculate the probability of positive and negative cascades, we can represent the model as a *biased random walk* on ΔN: the model begins with $\Delta N = 0$, and on each turn, an agent will either choose A (increasing ΔN by one) or choose B (decreasing ΔN by one). We continue to assume the simplified model with $\Delta P = 0$ and $k \gg 1$. Thus an agent's choice is made based on his preference Δu, his weight of public opinion q, and the current value of ΔN. Assuming as above that a player is randomly selected from the population each turn, the probability of increasing or decreasing ΔN can be calculated from x (the proportion of the population which prefers A), z (the proportion of the population which is highly dependent on public opinion), and the current ΔN. Let p^+, p^-, and p^0 denote the probabilities of moving right (i.e. increasing ΔN) when $\Delta N > 0$, $\Delta N < 0$, and $\Delta N = 0$ respectively. If $\Delta N > 0$, any agent who either prefers A or places high weight on public opinion will choose A; thus the probability of moving right is $p^+ = z + x(1 - z)$. If $\Delta N < 0$, an agent will only choose A if he prefers A and places low weight on public opinion; thus the probability of moving right is $p^- = x(1 - z)$. Finally, if $\Delta N = 0$, the agent will choose according to his private preference: the probability of increasing ΔN is $p^0 = x$. Figure 2 gives a pictorial representation of this random walk.

To analyze this random walk, we first consider the positive and negative regions as two separate random walks, then consider the

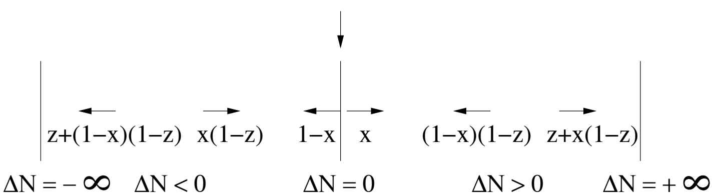

Figure 2. Random walk for sequential choice dependent on all previous actions

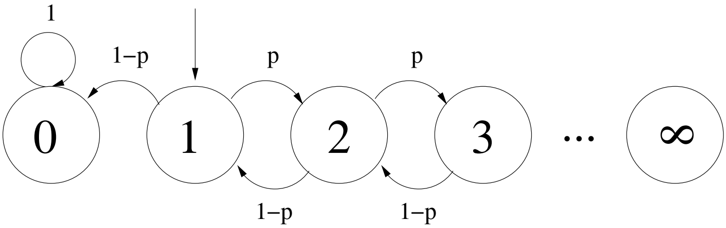

Figure 3. A directed random walk with absorbing boundary (Markov chain representation)

random walk formed by joining the two at the origin. Each region can be treated as a directed random walk with an absorbing boundary at the origin, and probability p of moving away from the origin (Figure 3). For the positive region, $p = p^+ = z + x(1 - z)$. For the negative region, $p = 1 - p^- = z + (1 - x)(1 - z)$. The behavior of each random walk depends on its value of p: if $p \leq .5$, the random walk will return to the origin in a finite number of time steps, but if $p > .5$, there is some non-zero probability of diverging to infinity (i.e. never returning to the origin). This probability p_∞ can be computed in terms of p: $p_\infty = 1 - (1 - p)/(p) = 2 - 1/p$ (see Shiryayev 1984, p. 547–549, for proof and further discussion).

We can now calculate the probability of positive and negative cascades by combining the two random walks. Since each random walk will either diverge to infinity or return to zero, we can represent this as a Markov chain with five states ($-\infty$, -1, 0, $+1$, $+\infty$). This Markov chain is shown in Figure 4. The $\pm\infty$ states are absorbing, and the 0 state will transition to either $+1$ or -1 with probabilities p^0 and $1 - p^0$ respectively. The ± 1 states will transition to $\pm\infty$ with probabilities $p_{+\infty} = \max(2 - (1)/(p^+), 0)$ and $p_{-\infty} = \max(2 - (1)/(1 - p^-), 0)$ respectively, and transition back to 0 otherwise. Thus if either $p^+ > .5$ or $p^- < .5$, the process will (with probability 1) end in one of the absorbing states. We note that for any $z > 0$, $p^+ > .5$, so divergence to $+\infty$ can occur. In this case, the proportion of agents choosing A is equal to $p^+ = z + x(1 - z) > x$, so we have a positive cascade effect. If $p^- < .5$, then divergence to $-\infty$ is also possible. In this case, the proportion of agents choosing A is equal to $p^- = x(1 - z) < x$, so we have a negative cascade effect. Moreover, since the proportion of

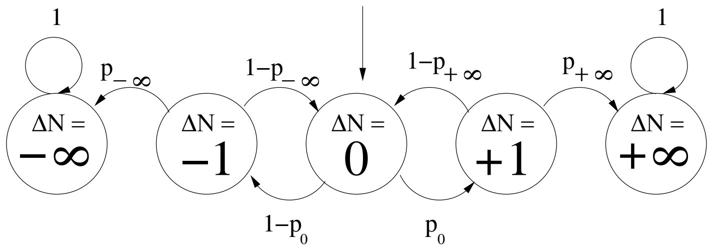

Figure 4. Simplified Markov chain for sequential random walk dependent on all previous actions

agents choosing A is less than .5, the process has converged to an inefficient norm in which the majority chooses a dispreferred action.

If $p_{+\infty} > 0$ and $p_{-\infty} > 0$, we can calculate the probability of a negative cascade as $((1 - p^0)p_{-\infty})/((1 - p^0)p_{-\infty} + p^0 p_{+\infty})$. Since this quantity is positive whenever $x(1 - z) < .5$ and $x < 1$, these are the *necessary conditions* for a negative cascade to occur. In other words, if at least half of the population are 'unconditional' followers of the majority action, then negative cascades cannot occur; if more than half the population either prefers the minority action or conditionally prefers the majority action, then negative cascades can occur. This means that, if at least half of a population of naïve Bayesian norm followers places high weight on public opinion ($z > .5$), negative cascades are always possible regardless of the size of the majority. Moreover, negative cascades can also occur when less than half place high weight on public opinion, as long as there is a sufficiently large minority.

3.2 Examples

Let us consider a population where a 75% majority prefers Coke to Pepsi ($x = .75$). We consider a sequential choice model as above: each individual in the population must make a choice between the two drinks, given only the aggregate result of all previous choices (the number of people, so far, who have chosen Coke and Pepsi respectively). As before, we assume that every individual has equal priors ($\Delta P = 0$) and a high weight of observation ($k \gg 1$).

First, we consider the case where 80% of the population puts high weight on popular opinion ($z = .8$). Based on the random walk model above, we compute $p^+ = z + x(1 - z) = .95$, $p^- = x(1 - z) = .15$, and $p^0 = x = .75$. Since $p^+ > .5$ and $p^- < .5$, both positive and negative cascades can occur. To compute the probability of a negative cascade, we first compute $p_{+\infty} = 2 - (1)/(p^+) = .947$ and $p_{-\infty} = 2 - (1)/(1 - p^-) = .824$. Then the chance that a negative cascade will occur is $((1 - p^0)p_{-\infty})/((1 - p^0)p_{-\infty} + p^0 p_{+\infty}) \approx 23\%$. In this case, $1 - p^- = 85\%$ of the population will drink Pepsi, even though only 25% prefer Pepsi; the other 60% prefer Coke, but believe (incorrectly) that they are following the majority preference. If a negative cascade does not occur, a positive cascade will result in $p^+ = 95\%$ of the population drinking Coke (75% because they prefer Coke, and 20% following the norm).

Second, we consider the case where 40% of the population puts high weight on popular opinion ($z = .4$). Using the random walk model, we calculate $p^+ = z + x(1 - z) = .85$, $p^- = x(1 - z) = .45$, and $p^0 = x = .75$. Since $p^- < .5$, a negative cascade can occur, even though the majority of the population does not put high weight on popular opinion ($z < .5$). We compute $p_{+\infty} = .824$ and $p_{-\infty} = .182$, and thus the probability of a negative cascade is $((1 - p^0)p_{-\infty})/((1 - p^0)p_{-\infty} + p^0 p_{+\infty}) \approx 6.9\%$. If a negative cascade occurs, $1 - p^- = 55\%$ of the population will drink Pepsi (25% because they prefer Pepsi, and 30% because they believe they are following the majority preference). If a negative cascade does not occur, a positive cascade will result in $p^+ = 85\%$ of the population drinking Coke (75% Coke-preferring, and 10% norm-following).

Finally, we consider the case where 20% of the population puts high weight on popular opinion ($z = .2$). Using the random walk model, we calculate $p^+ = .8$, $p^- = .6$, and $p^0 = .75$. Since p^- is not less than .5, we know that negative cascades cannot occur, and a positive cascade always occurs. As a result, $p^+ = 80\%$ of the population will choose Coke.

4. Generalizing the Model

At this point we reconsider some of the simplifying assumptions made in our model (constant and preference-independent parameters, equal priors, and high weight of observation). These assumptions may not hold in certain real-world interactions, so we consider the effects of relaxing each assumption on the large-scale results of our model (i.e. the frequency and magnitude of positive and negative cascades). In Sections 4.1–4.4, we maintain the assumptions that each individual puts either very high weight or very low weight on popular opinion, and similarly for weight of observation. In Sections 4.5 and 4.6, we allow for continuous values of these parameters, ranging from very high to very low. This generalization complicates the analysis of the random walk model, since the transition probabilities become dependent on N (the total number of choices) as well as ΔN (the difference in the observed numbers of A and B choices). As a result, we use simulation rather than obtaining an exact solution; nevertheless, we can obtain results to any desired degree of accuracy by repeated simulation. We demonstrate that allowing for continuous parameter values does not have a significant impact on the general results of our model, thus justifying our decision to focus on the variety of interesting cases where an exact closed-form solution may be obtained.

4.1 Preference-Dependent Parameters and Fanaticism

Our first assumption in the simplified model was that each individual's parameters (weight of public opinion q and weight of observation k) are independent of his preference for either A or B. But in many circumstances, different groups may rely more or less strongly on their observations of others' behavior. For example, a group with strongly pacifistic beliefs may refuse to go to war even if an overwhelming majority supported the war, while many supporters of the war might reconsider their stance if it was almost universally opposed. Thus we may have a situation in which individuals' parameters may be highly *correlated* with their preference Δu. Here we focus on one such effect, which we term *fanaticism*. A segment of the population is 'more fanatical' than another segment of the population if it holds more strongly to its private preferences and priors, and relies less on observations of others' behavior. Most importantly, then, a more fanatical group will have lower weights of public opinion q than a less fanatical group; additionally, the more fanatical group may have lower weights of observation k, along with strongly biased priors ΔP. Here we focus on the simpler case where all individuals have equal priors and high weight of observation; however, we allow different probabilities of depending on public opinion, conditioned on whether the individual prefers A or B. Let z_A be the probability that an individual is dependent on public opinion given that he prefers A, and let z_B be the probability that an individual is dependent on public opinion given that he prefers B. We can calculate the probability of positive and negative cascades using the random walk model as before, except that z is different for $\Delta N > 0$ and $\Delta N < 0$. For $\Delta N > 0$, an individual will choose action A if he prefers action A, or if he prefers action B but is dependent on public opinion: thus $p^+ = z_B + x(1 - z_B)$. Similarly, for $\Delta N < 0$, an individual will choose action A if he prefers action A and is not dependent on public opinion: thus $p^- = x(1 - z_A)$. Finally, for $\Delta N = 0$, an individual will choose his private preference as before: thus $p^0 = x$. We can calculate the probability of positive and negative cascades from p^+, p^-, and p^0 as above.

If $z_A > z_B$, this means that those preferring the minority action B are more 'fanatical' and less likely to change their beliefs based on public opinion. This will result in a higher probability of negative cascades and smaller probability of positive cascades. Also, if a negative cascade occurs, its magnitude (i.e. the proportion choosing B) will be increased, but if a positive cascade occurs, its magnitude (i.e. the proportion choosing A) will be decreased.

We reconsider the above example in which 75% of the population prefers Coke, and 40% of the population is dependent on public opinion ($x = .75$, $z = .4$). However, we assume that the influence of public opinion is different for Coke and Pepsi drinkers: 50% of Coke drinkers and 10% of Pepsi drinkers are dependent on public opinion ($z_A = .5$, $z_B = .1$). From this information we calculate $p^+ = z_B + x(1 - z_B) = .775$ and $p^- = x(1 - z_A) = .375$; note that $p^0 = x = .75$ as before. This gives us a probability of negative cascades of 16%, as compared to 7% for the symmetric case $z_A = z_B = z$. Also, if a negative cascade occurs, $1 - p^- = 63\%$ of the population will choose B, as opposed to 55% for the symmetric case; if a positive cascade occurs, $p^+ = 78\%$ of the population will choose A, as opposed to 85% for the symmetric case.

If $z_B > z_A$, this means that those preferring the minority action B are more likely to be swayed by the observed preference of the majority: in this case, the minority is *less* fanatical in their beliefs than the majority. As a result, the probability and magnitude of a negative cascade will be decreased, and the probability and magnitude of a positive cascade will be increased; this may even prevent negative cascades from occurring altogether. If in the above example, 70% of Pepsi drinkers and 30% of Coke drinkers were dependent on public opinion, we would have $p^+ = z_B + x(1 - z_B) = .925$ and $p^- = x(1 - z_A) = .525$. As a result, negative cascades would not occur, and we would have a population where 93% of individuals drink Coke (75% Coke-preferring and 18% norm-following), as opposed to 85% in the symmetric case.

4.2 Unequal Priors; Viral and Mass Marketing

Our second assumption in the simplified model was that all individuals have equal priors: if they have not observed any actions, they assume that the majority preference is equally likely to be *A* or *B*. However, unequal priors can result in many circumstances from an individual's beliefs, prejudices, and past experiences. For example, advertisements or propaganda may attempt to convince the population that a particular product is majority-preferred, and vocal expressions of support for (or protest against) a candidate or policy may lead others to (potentially inaccurate) estimates of its base of support.

Manipulation of the prior probabilities through advertisement or public dissemination of information may dramatically affect the probability and magnitude of cascade effects, especially if some fraction of the population has low weight of observation (and thus, depends heavily on their priors in making a decision). We consider low weights of observation in the following subsection; for now, we continue to assume that all individuals have high weight of observation ($k \gg 1$). In this case, any difference in the observed counts of actions *A* and *B* will outweigh the influence of the prior, thus p^+ and p^- are the same as in the original model. The only change is in p^0, the probability that an individual chooses *A* when the number of observations of *A* and *B* are equal. When $\Delta N = 0$, an individual will choose according to his prior if his weight of public opinion is high, and according to his private preference if his weight of public opinion is low. We assume that every member of the population has a prior weighted toward either *A* or *B*; let y be the proportion of the population with priors weighted toward *A*. Thus if $\Delta N = 0$, an individual will choose *A* with probability $p^0 = yz + x(1 - z) = x + (y - x)z$. Since only p^0 is changed, unequal priors (assuming high weight of observation) will change the probabilities but not the magnitudes of positive and negative cascades. If $y > x$, then the number of people believing that the majority is A is greater than the number of people actually preferring A, and the probability of negative cascades is decreased. If $y < x$, then the number of people believing that the majority is B is greater than the minority of people actually preferring B, and the probability of negative cascades is increased.

Reconsider the example with 75% of the population preferring Coke, and 40% of the population placing high weight on public opinion ($x = .75$, $z = .4$). We consider three cases: one where everyone initially believes (correctly) that Coke is majority-preferred, one where everyone initially believes (incorrectly) that Pepsi is majority-preferred, and one where the beliefs are split equally. For all examples, we calculate $p^+ = .85$ and $p^- = .45$ as above. If everyone believes a priori that Coke is preferred, we have $y = 1$, and thus $p^0 = x + (y - x)z = .85$. In this case, the probability of negative cascades would be only 3.8% (as opposed to 6.9% with equal priors). If everyone believes a priori that Pepsi is preferred, we have $y = 0$, and thus $p^0 = x + (y - x)z = .45$. In this case, the probability of negative cascades is increased to 21%. Finally, if half the population believes a priori that Coke is preferred, we have $y = .5$, and thus $p^0 = x + (y - x)z = .65$. In this case, the probability of negative cascades is 11%. Note that the probability of negative cascades is higher if the population is evenly divided with respect to priors, than if all members of the population have uninformative priors.

The ability to change the proportions of positive and negative cascades through manipulation of priors is extremely important in fields such as marketing and politics. In the example given above, a successful mass marketing campaign by Pepsi would triple the probability that Pepsi-drinking is adopted as a norm, while a successful mass marketing campaign by Coke would halve that probability. In fact, the impact of advertising is probably more significant than this since some fraction of the population might have low weight of observation; we consider this in the following subsection.[7] We also note that dissemination of public information by a reputable source (e.g. the government, or trusted news media sources) could decrease the probability of negative cascades significantly by revealing the true majority preference: this would not only change people's priors in the direction of the true majority, but also decrease the weight of observation (reflecting people's increased confidence in their prior estimates). This corresponds to the observations of Bikhchandani et al. (1992) and others that cascade effects are *fragile*; since the information content of a cascade is low, a negative cascade can be overturned by the public release of very little new information. For example, a news agency might release a public opinion poll based on the (private) preferences of a sufficiently large sample of the population, or the government might provide health warnings or other items which give clear indications of which norms are 'efficient' or 'inefficient.'

Finally, we consider the role that this model might play in schemes for 'viral marketing.' The goal of viral marketing is to 'take advantage of networks of influence among customers to inexpensively achieve large changes in behavior' (Domingos and Richardson 2001). The idea is that, rather than using company resources to market to a large number of potential customers, we market to only a smaller number of customers who are most 'influential' (able to communicate their preference to a large number of others). In a sequential choice scenario, it is clear that the most influential decision-maker is the first person to make a decision: thus if a viral marketer knew the order of decisions, and could only afford to market to a single customer, his optimal strategy would be to market to the first decision-maker. This is why, for instance, presidential candidates spend a significant amount of time and energy campaigning in the state of New Hampshire, even though the state has relatively few electoral votes: they hope that winning the first primary will create a 'bandwagon' effect and increase the probability of victory in succeeding primaries.

Let us consider the Coke-Pepsi example above, with 75% of individuals preferring Coke and 40% of individuals placing high weight on public opinion ($x = .75$, $z = .4$). We assume that all individuals have high weight of observation. Furthermore, let us assume that marketing a product to an individual biases his priors toward that product (i.e. convinces him that the product is majority-preferred, unless his observations indicate otherwise); all other individuals have uniform priors. Thus if Coke is marketed to the first individual, he will choose Coke if he either prefers Coke or has high weight of public opinion, i.e. with probability $z + x(1 - z) = .85$. Thus the random walk on ΔN has probability $.85p_{+\infty} = .7$ of diverging to $+\infty$ before its first return to zero, probability $.15p_{-\infty} = .027$ of diverging to $-\infty$ before its first return to zero, and probability $1 - (.7 + .027) = .273$ of returning to zero at least once. If it returns to zero, then the probability of a negative cascade is 6.9% as above.

From this, we find that the probability of a negative cascade (i.e. the majority of the population drinking Pepsi) is $.027 + (.273)(.069) = 4.6\%$, as compared to 6.9% if no marketing was done, and 3.8% if Coke was marketed to all customers. It is clear from this example that viral marketing is extremely successful in the sequential choice case; in this case it achieved 74% of the impact of mass marketing at a tiny fraction of the cost.

If Pepsi is marketed to the first individual, he will choose Coke only if he both prefers Coke and has low weight of public opinion, i.e. with probability $x(1 - z) = .45$. Thus the random walk on ΔN has probability $.45p_{+\infty} = .371$ of diverging to $+\infty$ before its first return to zero, probability $.55p_{-\infty} = .1$ of diverging to $-\infty$ before its first return to zero, and probability $1 - (.371 + .1) = .529$ of returning to zero at least once. From this, we find that the probability of a negative cascade is $.1 + (.529)(.069) = 14\%$, as compared to 6.9% if no marketing was done, and 21% if Pepsi was marketed to all customers. Again, the example demonstrates that viral marketing was able to achieve a significant impact (48% of the impact of mass marketing) with very little cost.

4.3 Low Weight of Observation

Our third assumption in the simplified model was that all individuals have high weight of observation: they tend to rely on empirical observations of others' actions rather than their prior beliefs about others' preferences. This assumption is fairly realistic, in that people tend to reject prior assumptions when these assumptions conflict with direct observations. In fact, Kahneman and Tversky (1973, 1974) argue that people tend to systematically underweight or ignore the influence of prior probabilities when making a judgment, instead focusing on the 'representativeness' of an example (its empirical similarity to a particular population). This does not imply, however, that all individuals have weight of observation so high that they ignore their priors. When the number of observations N is small, or there is not a strong observed majority preference (i.e. $|\Delta N|$ is small compared to N), it makes sense for a rational decision-maker to rely more heavily on his prior information. This is particularly important when the prior is obtained from a reliable source (e.g. government news releases) and when empirical observations may be heavily biased by cascade effects. Moreover, Anderson and Holt (1997) find no evidence of a representativeness bias in their experimental results on 'informational cascades in the laboratory': though there were a significant number of errors, subjects tended to make a decision consistent with Bayes' rule even when it conflicted with representativeness. As they state, the representativeness heuristic may not apply when the priors in the model are based on the subjects' own inferences, rather than being given in the experimental instructions.

We find that lowering the weight of observation has two main effects. First, it tends to reduce the probability p^+ and increase the probability p^-, which has the effect of reducing the magnitude of cascades, and may also significantly change the relative probabilities of positive and negative cascades. Second, it amplifies the effects of biased prior probabilities: if priors are strongly biased toward the majority-dispreferred action, this may even result in a situation where only negative cascades (i.e. no positive cascades) can occur.

We consider the case where all individuals have priors biased either toward A or toward B. As above, let y be the proportion of the population with priors biased toward A; also let x be the proportion preferring A, and z be the proportion placing high weight on public opinion. Finally, we assume that all individuals have either very high weight of observation ($k \gg 1$) or very low weight of observation ($k \approx 0$): let w be the proportion with high weight of observation. For simplicity, we assume that priors, weight of public opinion, and weight of observation are uncorrelated and preference-independent. In this case, we can once again use the random walk model, and condition the probability of moving right on the current value of ΔN. In all cases, an individual will choose A if he prefers A and places low weight on public opinion. Similarly, an individual will always choose A if his priors are biased toward A and he has high weight of public opinion and low weight of observation. If an individual has high weights of public opinion and observation, he will choose A if $\Delta N > 0$, B if $\Delta N < 0$, and according to his prior if $\Delta N = 0$. From this analysis, we calculate $p^+ = x(1-z) + y(1-w)z + wz$, $p^- = x(1-z) + y(1-w)z$, and $p^0 = x(1-z) + y(1-w)z + ywz$. Simplifying, we obtain $p^0 = x + (y-x)z$ as in the previous subsection; thus the weight of observation does not affect choices when $\Delta N = 0$. We also obtain $p^+ = z + x(1-z) - (1-y)(1-w)z \le z + x(1-z)$, and thus p^+ will be decreased by lowering the weight of observation whenever $y < 1$ and $z > 0$. Similarly, we note that $p^- = x(1-z) + y(1-w)z \ge x(1-z)$, and thus p^- will be increased by lowering the weight of observation whenever $y > 0$ and $z > 0$. Thus, if the population is divided with respect to priors ($0 < y < 1$) and some fraction has high weight of public opinion ($z > 0$), the magnitudes of both positive and negative cascades will be reduced. Reducing p^+ and increasing p^- may also affect the relative probabilities of positive and negative cascades. If $y \approx 1$, p^- will be increased significantly more than p^+ is reduced, and thus the probability of negative cascades will be reduced. If $y \approx 0$, on the other hand, p^+ will be reduced significantly more than p^- is increased, and thus the probability of negative cascades will be increased. It is even possible to have situations where negative cascades are guaranteed; this will occur if $p^+ < .5$ (since $p^- \leq p^+$, we will also have $p^- < .5$ in this case).

To demonstrate these effects, we reconsider the example where 75% of the population prefers Coke ($x = .75$) and 40% of the population places high weight on public opinion ($z = .4$). Again we consider three cases: $y = 1$ (everyone believes a priori that Coke is preferred), $y = 0$ (everyone believes a priori that Pepsi is preferred), and $y = .5$ (prior beliefs are evenly divided between Coke and Pepsi). This time, however, we assume that the majority (90%) of the population has low weight of observation: let $w = .1$. For $y = 1$ we calculate $p^+ = p^0 = .85$ and $p^- = .81$: thus a positive cascade will always occur, causing 85% of the population to drink Coke. Compare this to the case in which all individuals had high weight of observation ($w = 1$), which still allowed for negative cascades to occur (with 3.8% probability). In this case, negative cascades can only occur when $w > \frac{7}{8}$. For $y = 0$, on the other hand, we calculate $p^+ = .49$ and $p^- = p^0 = .45$: thus a negative cascade will always occur, causing 55% of the population to drink Pepsi. This is very different than the case of $w = 1$, where the probability of negative cascades was only 21%. In this case, positive cascades can only occur when $w > \frac{1}{8}$. Finally, for $y = .5$ we calculate $p^+ = .67$, $p^0 = .65$, and $p^- = .63$. Thus we have a situation where over half (67%) of the population chooses the majority preference, but surprisingly, the proportion of individuals choosing the majority preference is actually *less* than the proportion of the majority! Even though ΔN goes to $+\infty$, and almost every individual who has high weights of public opinion and observation chooses A, this is outweighed by the large number of individuals who choose B (even though they prefer A) because of their incorrect prior beliefs that B is majority-preferred. This somewhat paradoxical result demonstrates that priors can have a significant effect on the results of collective decisions, especially when the weight of observation is low.

4.4 Varying Weight of Public Opinion

Next we examine the assumption that an individual's weight of public opinion q is constant, rather than varying with the number of observations N. Relaxing this assumption allows us to connect our model directly to the prior work of Bikhchandani et al. (1992), a framework which is also used in the experiments of Anderson and Holt (1997) and is closely related to the models of Arthur (1989) and Bicchieri and Fukui (1999).

In particular, we consider situations in which one of the two actions (A or B) is 'objectively correct,' and would be preferred by all individuals if they had complete information. This occurs, for example, in the laboratory experiments of Anderson and Holt (1997), where each individual in turn must guess from which urn a ball has been drawn. Similarly, the reputational cascade model of Bicchieri and Fukui (1999) assumes that every individual wants to choose the action which is preferred by the majority, regardless of his own private preference. In each of these cases, an individual values his own information or preference no more than that of any other individual; thus the 'rational' decision is to follow the majority of observed decisions, counting the individual's private information as a single observed decision (Bikhchandani et al, 1992). This can be expressed as a special case of our model where we set $q = N$, and make the assumptions of equal priors and high weight of observation. An individual will choose A in our model if $\Delta u + q(\Delta P + k\Delta N)/(1 + kN) > 0$. Setting $\Delta P = 0$, $k \gg 1$, and $q = N$, this simplifies to $\Delta u + N(\Delta N/N) = \Delta u + \Delta N > 0$. Since we assumed (without loss of generality) that $|\Delta u| = 1$, an individual who prefers A will choose A if $\Delta N > -1$, and an individual who prefers B will choose A if $\Delta N > 1$. For k large but finite, the weight of public opinion is actually slightly less than one $(kN)/(kN + 1)$ so ties (i.e. when $\Delta u + \Delta N = 0$) will be broken by the individual's private preference. For $k = \infty$, the weight of opinion is exactly one, so an individual preferring A when $\Delta N = -1$ or preferring B when $\Delta N = 1$ is indifferent between the two actions. In this case, as in the original model of Bikhchandani et al, we assume that he flips a coin to decide.

We should note that the assumption of $q \propto N$ applies to *strong cascade* situations, when an individual considers his preference (or private information) to be as relevant as the decisions of q_0 other individuals; in this case, we set $q = N/q_0$. On the other hand, the assumption of q constant applies to *weak cascade* situations, when an individual weights his own preference proportional to the average preference of the population as a whole. As discussed above in Section 1.2, the former is a better assumption when describing problems with an objectively correct decision that must be inferred from private information and observations, while the latter is a better assumption when describing phenomena such as fashions, fads, and customs, in which there is no objectively correct decision. In deciding whether to follow an observed custom, for instance, a rational individual must decide based on the strength of his own personal preference as well as his estimate of the prevalence and importance of that custom within society; a decision to reject a behavioral norm may be rational for a given individual even if that norm is strongly entrenched and the individual risks public disapproval by doing so.

Let us first consider the model of Arthur (1989), where agents must choose sequentially between two technologies A and B, and the payoffs for adopting a technology are increased proportional to the number of previous adopters of that technology. Arthur assumes two groups with distinct preferences: type R individuals have payoffs $a_R + rN_A$ for choosing A and $b_R + rN_B$ for choosing B, and type S individuals have payoffs $a_S + sN_A$ for choosing A and $b_S + sN_B$ for choosing B. Arthur assumes $a_R > b_R$, $a_S < b_S$, and an equal proportion of types R and S. Thus we have three distinct regimes: for $\Delta N > (b_S - a_S)/(s)$ both types will choose A, for $\Delta N < (b_R - a_R)/(r)$ both types will choose B, and otherwise each individual will choose his private preference (A for R-types, B for S-types). Thus the probabilities of increasing ΔN are 1 for $\Delta N > (b_S - a_S)/(s)$, 0 for $\Delta N < (b_R - a_R)/(r)$, and $\frac{1}{2}$ otherwise, and the model simplifies to an unbiased random walk with absorbing boundaries. This is a simple case of our model (which is, more generally, a biased random walk with transition probabilities dependent on ΔN and N); it can be easily represented by setting $\Delta P = 0$, $k \gg 1$, $x = .5$, and $q = N/q_0$, where $q_0 = q_{0,R} = (a_R - b_R)/(r)$ for R-types and $q_0 = q_{0,S} = (b_S - a_S)/(s)$ for S-types. Then the probability of A being locked-in can be easily calculated as $(q_{0,R})/(q_{0,R} + q_{0,S})$

for Arthur's model; we note that both A and B cascades are total (all individuals adopt the locked-in technology) and irreversible once the absorbing boundary has been reached.

Next, we consider how the Coke/Pepsi example would be treated by the Bikhchandani et al. model. Let us assume as above that 75% of the population prefers Coke ($x = .75$). In order to calculate the probability of positive and negative cascades, we must consider five distinct regions. First, whenever $\Delta N < -1$, an individual will always choose Pepsi, and thus a negative cascade results. Second, whenever $\Delta N > 1$, an individual will always choose Coke, and thus a positive cascade results. If $\Delta N = 0$, an individual will choose according to his private preference; assuming that individuals are picked from the population randomly, this implies that the probability of moving right is $x = .75$. If $\Delta N = -1$, an individual will choose Pepsi if he prefers Pepsi, and flip a coin otherwise; this implies that the probability of moving right is $x/2 = .375$. If $\Delta N = 1$, an individual will choose Coke if he prefers Coke, and flip a coin otherwise; thus the probability of moving right is $x + (1 - x)/(2) = .875$. From these values, we find that the probability of negative cascades is $(.25(1 - .375))/(.25(1 - .375) + .75(.875)) = 19\%$. We also note that, since the $\Delta N > 1$ and $\Delta N < -1$ states are absorbing, it often takes only two decisions to create a irreversible positive or negative cascade. Our model, on the other hand, allows cascades to be reversed after any number of initial decisions, though the probability of reversing a cascade would be very small if $|\Delta N| \gg 0$. Also, the Bikhchandani et al. model (like Arthur's increasing returns model) results in *total* cascades, i.e. the proportion of the population choosing an action goes to either 1 or 0 as the number of choices N goes to infinity. Our model, by allowing some individuals to have lower weight of public opinion, also allows for *partial* cascades in which some individuals do not choose the majority-chosen action. This is a more realistic model of preference decisions, in which non-homogeneous results (such as part of the population choosing Coke, and part choosing Pepsi) are extremely common. Thus our model is clearly a generalization of previous models of cascade effects, allowing for more realistic populations with varying priors, weights of public opinion, and weights of observation.

4.5 Simulating the General Model

We now consider the macromodel in its full generality, where we allow for continuous distributions of individuals' priors ΔP, weights of public opinion q, and weights of observation k. We continue to assume (without loss of generality) that $|\Delta u| = 1$ for all individuals; let $x \geq .5$ be the proportion of individuals preferring A (i.e. individuals with $\Delta u = 1$). In the general case, as discussed in Section 3, we have a biased random walk on ΔN, where N is an additional state variable that affects the transition probabilities. For a given state $(\Delta N, N)$, we transition either to state $(\Delta N + 1, N + 1)$ or state $(\Delta N - 1, N + 1)$ with probabilities p and $1 - p$ respectively, where:

$$p = \Pr\left(\Delta u + q\frac{\Delta P + k\Delta N}{1 + kN} > 0\right)$$

$$= x\Pr\left(q\frac{\Delta P + k\Delta N}{1 + kN} > -1\right) + (1 - x)\Pr\left(q\frac{\Delta P + k\Delta N}{1 + kN} > 1\right)$$

For arbitrary distributions of ΔP, q, and k, no general closed-form solution exists for this random walk model. However, it is simple to simulate the model, assuming a large but finite population of M individuals (here we use $M = 10000$), and known distributions of ΔP, q, and k. On each time step $t = 1 \ldots M$, we simulate the choice of one individual by sampling Δu, ΔP, q, and k from their respective distributions, and updating ΔN and N accordingly. Once all M individuals have made their choices, we calculate the proportion of individuals x_{obs} choosing A, and compare this to x: if x_{obs} is significantly greater than x, a positive cascade has occurred, and if x_{obs} is significantly less than x, a negative cascade has occurred. Our significance test is simple, since choices are binomially distributed under the null hypothesis of no cascade: we assume that a cascade has occurred whenever x_{obs} is outside the 95% confidence interval for x, i.e. when $|x_{obs} - x| > 1.96\sqrt{(x(1 - x)/M)}$. We repeat the entire simulation 10000 times, allowing us to calculate the probabilities of positive and negative cascades, as well as the average magnitude of each type of cascade.

For our first set of simulations, we continue to make the simplifying assumptions of $\Delta P = 0$ and $k \gg 1$, but we allow q to vary uniformly over the interval $[0, q_{max}]$. We test for two values of x, a slight majority ($x = .55$) and a larger majority ($x = .75$), and use values of q_{max} ranging from 1 to 50. See Table 1 for results: P_{pos}

Table 1. Probabilities and magnitudes of positive and negative cascades, $k \gg 1$

	$x = .55$				$x = .75$			
q_{max}	P_{pos}	x_{pos}	P_{neg}	x_{neg}	P_{pos}	x_{pos}	P_{neg}	x_{neg}
1	.027	.562	.026	.538	.023	.760	.026	.740
2	.025	.562	.022	.538	.711	.785	.011	.740
3	.170	.582	.040	.538	1.000	.892	.000	–
4	.491	.827	.258	.367	1.000	.924	.000	–
5	.548	.881	.382	.166	.952	.937	.044	.604
10	.554	.950	.446	.063	.776	.974	.224	.092
20	.549	.976	.451	.029	.764	.987	.236	.041
30	.552	.985	.449	.019	.753	.992	.247	.026
40	.558	.988	.442	.014	.747	.994	.253	.020
50	.548	.991	.452	.011	.757	.995	.243	.015

and P_{neg} are the probabilities of positive and negative cascades, and x_{pos} and x_{neg} are the average magnitudes of positive and negative cascades (i.e. the average proportion x_{obs} of individuals choosing the majority preference A in each case).

For both $x = .55$ and $x = .75$, we can divide the results into three distinct regimes. For sufficiently low values of q_{max}, we find that *no cascades occur*, and we have $x_{obs} \approx x$. More precisely, the results in this case are indistinguishable from those generated by a binomial distribution with parameter x: x_{obs} falls outside of the 95% confidence interval for x only 5% of the time, with approximately equal probabilities of $x_{obs} < x$ and $x_{obs} > x$, and thus the effects of cascades are negligible. For high values of q_{max}, we find that *near-total cascades occur*: if a positive cascade occurs then nearly everyone adopts the majority preference ($x_{obs} \approx 1$), and if a negative cascade occurs then nearly everyone adopts the minority preference ($x_{obs} \approx 0$). Also, in this case we find that cascades occur with probability near 1, with the probability of positive cascades approximately equal to the proportion of the majority x. This is not surprising, since in cases where the weight of public opinion is high, nearly everyone will follow the lead of the first decision-maker, who will choose A with probability x. For moderate values of q_{max}, we find that *partial cascades occur*: if a positive cascade occurs we have $x < x_{obs} < 1$, and if a negative cascade occurs we have $0 < x_{obs} < x$. The magnitudes of both positive and negative cascades increase with q_{max}, i.e. positive cascades have higher x_{obs} and negative cascades have lower x_{obs} as q_{max} increases. Also, the relative probability of positive cascades as compared to negative cascades $\left(P_{pos}/P_{neg}\right)$ is highest for moderate values of q_{max}. For $x = .55$, both P_{pos} and P_{neg} increase with q_{max}, but P_{pos} increases more rapidly: for example, for $q_{max} = 4$, we have $P_{pos} = .491$ (nearly equal to its limiting value of $x = .55$) while $P_{neg} = .258$ (only slightly more than half of its limiting value of $1 - x = .45$). Similarly, for $x = .75$, P_{pos} increases much more rapidly than P_{neg}, and this results in positive cascades occurring with probability significantly greater than x. In fact, for $q_{max} = 3 \ldots 4$, we find that positive cascades always occur, and negative cascades never occur, while for $q_{max} = 5$, positive cascades occur 95% of the time.

It should be noted that these results are very similar to those obtained in the case where all individuals have either very high or very low weights of public opinion. For z (the proportion of individuals with high weights of public opinion) near 1 in the simplified model, we have $p^+ \approx 1$ and $p^- \approx 0$, so both positive and negative cascades are near-total cascades, and the probability of positive cascades is approximately $p^0 = x$. For z sufficiently small in the simplified model, we have a slight positive cascade where only those individuals with very high weights of public opinion are affected by the cascade; as z approaches zero, the proportion choosing A converges to $p^+ = z + x(1 - z) \approx x$. Finally, for moderate values of z, we have partial cascades, where the probability of positive cascades is larger than x, and also the magnitude of positive cascades is larger than the magnitude of negative cascades (i.e. $x_{pos} > 1 - x_{neg}$). All of these results are also visible in Table 1; thus the simplified case (with all weights of public opinion q either very high or very low) displays all of the interesting behaviors of the general case with a continuous distribution of q.

For our second set of simulations, in addition to allowing q to vary uniformly over $[0, q_{max}]$, we also allow k to vary uniformly over $[0, k_{max}]$, and allow ΔP to vary uniformly over $[-1, 1]$. For simplicity, we fix $q_{max} = 10$; we again test for two values of x ($x = .55$ and $x = .75$), and use values of k_{max} ranging from 0 to 100. See Table 2 for results.

We first consider the results for large values of k_{max}, given $q_{max} = 10$ and the uniform distribution of priors. In this case, we find that the magnitudes of positive and negative cascades are identical to those for the equivalent case with no priors (i.e. $q_{max} = 10$,

Table 2. Probabilities and magnitudes of positive and negative cascades, $q_{max} = 10$

	$x = .55$				$x = .75$			
k_{max}	P_{pos}	x_{pos}	P_{neg}	x_{neg}	P_{pos}	x_{pos}	P_{neg}	x_{neg}
0	.000	–	1.000	.516	.000	–	1.000	.582
.01	.930	.853	.047	.385	1.000	.948	.000	–
.02	.875	.888	.093	.197	1.000	.959	.000	–
.05	.784	.916	.189	.098	.999	.966	.001	.131
.1	.718	.930	.268	.079	.991	.969	.009	.111
.2	.664	.939	.326	.070	.951	.971	.049	.103
.5	.600	.945	.397	.066	.847	.972	.153	.096
1	.569	.948	.429	.064	.777	.973	.223	.094
10	.522	.950	.478	.063	.639	.974	.361	.092
100	.524	.950	.476	.063	.617	.974	.383	.092

$k \gg 1$, and $\Delta P = 0$ for all individuals). However, the probability of negative cascades is substantially higher for uniformly distributed priors than for no priors: for uniform priors, P_{neg} was 48% for $x = .55$ and 38% for $x = .75$, while these probabilities were reduced to 45% and 24% respectively for no priors. This corresponds to two of the observations made in Section 4.2: first, when weights of observation are sufficiently high, the probabilities but not the magnitudes of cascades are affected by biased priors. Second, the probability of negative cascades is higher for evenly divided priors than for no priors, since the proportion of individuals with priors biased toward the minority preference B (50%) is higher than the proportion of individuals actually preferring B.

Next we consider the effects of decreasing k_{max} on the probabilities and magnitudes of cascades. For lower values of k_{max}, the magnitudes of both positive and negative cascades are reduced; as discussed in Section 4.3, this effect occurs because a smaller proportion of individuals are influenced by their observations of others' behavior. Additionally, the probability of negative cascades decreases with decreasing k_{max}: P_{neg} decreases from 48% down to 5% for $x = .55$, and from 38% down to 0% for $x = .75$. This effect results because the increase in p^- from lower weights of observation k outweighs the decrease in p^+: in particular, negative cascades will not occur at all when $p^- \geq .5$, and this effect is visible for $x = .75$ and $k_{max} \leq .02$.

Finally, we note that in the special case of $k_{max} = 0$, individuals are not affected by their observations of others' behavior. As a result, each individual will choose *A* or *B* independently, depending on whether $\Delta u + q\Delta P$ is positive or negative. In this case, for priors equally distributed toward *A* and *B*, the proportion of individuals choosing *A* will be between .5 (the proportion whose priors are biased toward *A*) and *x* (the proportion who actually prefer *A*). Thus the net effect is equivalent to a negative cascade (i.e. a proportion less than *x* adopt *A*), but this effect results from independent choices with biased priors rather than any actual 'cascade' effect as such. We should also note that this effect will not occur for any positive value of k_{max}: if the population is sufficiently large, individuals' observations will eventually overwhelm their priors for any $k_{max} > 0$.

Thus we have used simulation to examine a variety of cases of the general model, allowing for continuous distributions of the parameters ΔP, q, and k. It is clear from our results that the effects observed in the general model are very similar to those obtained in the simplified model, where all individuals have either very high or very low weights of observation and public opinion. In particular, for either weight of public opinion q or weight of observation k, we can achieve equivalent effects by either increasing the mean of the continuous distribution of that parameter in the general model, or increasing the *proportion* of the population with high values of that parameter in the simplified model. These correspondences justify our decision to focus on the simplified model, allowing us to compute exact closed-form solutions for the magnitudes and probabilities of positive and negative cascades, while achieving the same range of cascade effects as in the more general model.

4.6 Simultaneous and Sequential Choice

At this point, we revisit one of the fundamental assumptions of our macromodel: the assumption of *sequential choice*. As in many of the previous models of cascade effects (e.g. Arthur 1989; Banerjee 1992; Bikhchandani et al. 1992), we assume that individuals make choices one at a time, and are influenced by the entire sequence of previously made decisions. Our assumption of sequentiality is somewhat weaker than that made by previous models, since individuals are not assumed to observe the entire sequence of previous choices, but only the aggregate counts N_A and N_B (the numbers of individuals choosing A and B respectively). Nevertheless, our assumption neglects several phenomena which are relevant to some of the real-world examples we consider. First, rather than observing the aggregate counts N_A and N_B, individuals may only be able to *sample* from this distribution, obtaining counts n_A and n_B based on the choices made in their immediate (spatial or temporal) vicinity. Models of this sort are considered in Arthur and Lane (1993) and Banerjee and Fudenberg (2004), and require somewhat different Bayesian updating rules than the present model. For the purposes of our discussion, we note that cascade effects (and in particular, negative cascades) can still occur when only a sample of previous choices are observed. In fact, we expect the probability of negative cascades to *increase,* since there is a greater probability that the sampled choices will not be representative of the decisions of society as a whole; see Bicchieri and Fukui (1999) for one example of this effect.

A second effect neglected in our model is that individuals may change their minds, *updating* a decision in light of subsequently made decisions. This effect may be treated in one of two ways: either individuals may be allowed to make multiple choices, each of which is counted separately, or individuals may be allowed to make a new choice which invalidates their previous choice. The former applies to examples such as product choice, where individuals may buy multiple products; this effect is already accounted for in our model, since we assume that individuals are chosen from the population with replacement. The latter applies to examples where only an individual's 'current' choice is visible, as may be the case in fashions or fads. In this case, the magnitudes of cascades would be increased, since early-choosing individuals who originally chose in opposition to the cascade may change their minds and decide to join the cascade as well.

The most important effect that our model has not yet considered is *simultaneity*: it is clear that in many cases, different individuals may make decisions simultaneously in different locations. One possible way of dealing with this effect is to assume that separate and independent cascades occur in different locations, with a strict sequence of decisions occurring for each distinct cascade. In this case, we must consider what happens when two cascades (which were formerly spatially isolated) meet: decision-makers at the intersection of the two cascades would be influenced by observations resulting from both cascades, and the resulting choices may propagate to regions formerly dominated by one cascade or the other. For a more detailed treatment of the spatial propagation of cascades, the reader is referred to the work of Watts (2002), who examines network models of cascade effects. We note, however, that even allowing for multiple spatially isolated cascades (with sequential decisions in each cascade) neglects the fact that essentially simultaneous decisions may in fact be made in the same spatial locale. Thus one simple but interesting extension to our model is to assume that decisions are made in *groups*: we assume T discrete time steps, where individuals are influenced by decisions made on previous time steps, but are not influenced by other decisions made on the current time step. The standard model of sequential choice is equivalent to assuming that exactly one individual makes a choice on any given time step; we can easily consider cases where G individuals make a choice simultaneously on a given time step, where the parameter G denotes the 'group size.' It is clear that cascade effects will still occur in the case where decisions are made in groups, as long as individuals both *observe* the decisions of previous groups, and are *influenced* by these decisions. It is also clear that the probability of negative cascades will be reduced if the size of early-deciding groups is large: negative cascades are much more likely to occur when a majority of the early decision-makers choose contrary to the majority preference of society as a whole, and (assuming that individuals are drawn at random from the population) this is very unlikely to occur when group size is large. Nevertheless, negative cascades can still occur, either because of sampling variance in small groups, or systematic bias in groups of any size.

To quantify these effects, we ran several simulations of the general model. As in our first set of simulations in Section 4.5, we made the simplifying assumptions of $\Delta P = 0$ and $k \gg 1$, but allowed q to vary uniformly over the interval $[0, q_{max}]$. However, rather than assuming that one individual chooses on each time step ($G = 1$), we assumed group sizes of $G = 10$ and $G = 100$; these results are given in Tables 3 and 4 respectively. In each case, we used two values of x ($x = .55$ and $x = .75$) as above. As expected, the most noticeable effect of increasing G was that the probability of positive cascades increased, and the probability of negative cascades decreased, with increasing group size. For $x = .55$ and large values of q_{max}, P_{neg} decreased from 45% for $G = 1$, to 34% for $G = 10$, to 11% for $G = 100$. For $x = .75$ and large values of q_{max}, P_{neg} decreased

Table 3. Probabilities and magnitudes of positive and negative cascades, group size 10

	$x = .55$				$x = .75$			
q_{max}	P_{pos}	x_{pos}	P_{neg}	x_{neg}	P_{pos}	x_{pos}	P_{neg}	x_{neg}
1	.026	.562	.024	.538	.024	.760	.025	.740
2	.023	.562	.026	.538	.669	.783	.011	.740
3	.060	.572	.025	.538	1.000	.892	.000	–
4	.279	.803	.053	.459	1.000	.926	.000	–
5	.439	.869	.105	.220	1.000	.943	.000	–
10	.694	.944	.298	.067	.988	.973	.012	.095
20	.665	.976	.335	.030	.980	.987	.020	.042
30	.656	.984	.344	.020	.979	.991	.021	.027
40	.658	.988	.342	.015	.980	.993	.020	.020
50	.664	.990	.336	.012	.978	.995	.022	.016

from 25% for $G = 1$, to 2% for $G = 10$, to near 0% for $G = 100$. For very high weights of public opinion q, negative cascades will only occur if a majority of the first group chooses the minority preference B; this phenomenon becomes increasingly uncommon as group size becomes large. As above, the probability of negative cascades may be even lower for moderate values of q_{max}: we note that for $q_{max} = 3 \ldots 4$ and $x = .75$, positive cascades always occurred, and negative cascades never occurred, regardless of group size. Another effect of increasing group size is that the transition from the 'no cascades' regime (low values of q_{max}) to the 'near-total cascades' regime (high values of q_{max}) occurs for larger threshold values of q_{max}. Thus for a given, moderate value of q_{max} (e.g. $q_{max} = 4$), we find that the magnitudes of both positive and negative cascades decrease with increasing group size. This may also prevent cascades from occurring for low values of q_{max}: for $x = .55$ and group size 100, non-negligible positive and negative cascades do not occur until $q_{max} > 3$ and $q_{max} > 5$ respectively, while for group size 1, these thresholds were $q_{max} > 2$ and $q_{max} > 3$ respectively.

Thus we have considered the assumption of sequential choice in our macromodel, and its influence on cascade effects. The most important result to note from our discussion is that strict sequential choice is not necessary for cascade effects to occur: as long as some individuals make decisions after other individuals, and these individuals both observe and are influenced by previous decisions,

Table 4. Probabilities and magnitudes of positive and negative cascades, group size 100

	$x = .55$				$x = .75$			
q_{max}	P_{pos}	x_{pos}	P_{neg}	x_{neg}	P_{pos}	x_{pos}	P_{neg}	x_{neg}
1	.026	.562	.026	.538	.024	.760	.027	.740
2	.026	.562	.024	.538	.544	.777	.017	.740
3	.026	.562	.026	.538	1.000	.887	.000	–
4	.035	.613	.025	.539	1.000	.922	.000	–
5	.116	.791	.025	.538	1.000	.939	.000	–
10	.886	.905	.021	.263	1.000	.970	.000	–
20	.936	.964	.064	.041	1.000	.984	.000	–
30	.910	.976	.090	.029	1.000	.989	.000	–
40	.894	.981	.106	.023	1.000	.991	.000	–
50	.889	.984	.111	.019	1.000	.992	.000	

cascades will occur. As discussed above, the size of these cascades is strongly influenced by the distribution of weights of public opinion, and cascade effects may range from negligible to total. It is also important to note the effects of simultaneous choice on the relative probabilities of positive and negative cascades: as the amount of simultaneity (parametrized by group size G) increases, negative cascades become increasingly rare, and this may even prevent negative cascades from occurring altogether.

5. Discussion: Revisiting the Micromodel

Lastly, we reconsider the underlying assumptions of our micromodel of naïve Bayesian norm followers. Given that cascade effects (and in particular, negative cascades) occur, it is possible that the choices an agent observes may not be an accurate reflection of the preferences of the population as a whole. More precisely, naïve Bayesian norm followers assume that all their observations are i.i.d. (independent and identically distributed) given the model. But if a cascade has occurred, the sequence of previous choices will be highly correlated. Why, then, do our rational decision-makers not take this possibility into account, and adjust their decisions accordingly? There are several possible answers to this question. First, the structure of the game is not assumed to be common knowledge: in fact, we typically assume that agents do not observe the *sequence* of previously made decisions, but only the aggregate numbers of *A* and *B* decisions made. Thus, agents may incorrectly assume that previous choices were made simultaneously rather than sequentially. This effect is very likely to occur when an individual can only observe the behavior of the others around him at a given point in time, and thus does not realize that the other agents' behavior emerged gradually through a process of imitation. If all previous choices were made simultaneously rather than sequentially, they would correctly reflect the private preferences (and prior beliefs) of a sample of the population, and cascade effects would not apply.[8]

A second cause of 'naïve' behavior is the phenomenon of *pluralistic ignorance*: individuals often have erroneous cognitive beliefs regarding the ideas, sentiments, and actions of others (O'Gorman 1975). In particular, individuals often believe that their private thoughts and feelings are different from those of others, even though their public behavior is identical (Miller and McFarland 1987). Bicchieri and Fukui (1999) discuss pluralistic ignorance in detail, and distinguish between two major classes of effects. First, pluralistic ignorance may result in the illusory belief that others hold a given set of values more strongly than the individual does himself: this effect is evident in the studies of O'Gorman (1975), where white Americans overestimated the support of other whites for racial segregation, and Prentice and Miller (1993), where college students tended to overestimate the prevalence of alcohol abuse among other students. The primary representation of this effect in our model is the weighting of individuals' priors ΔP; as discussed above, biased priors (especially when combined with low weights of observation) can lead to situations where incorrect beliefs cause a high probability of negative cascades.

More relevant to the current discussion, however, is Bicchieri and Fukui's second class of effects caused by pluralistic ignorance: even when individuals correctly identify the positive values of a group, they underestimate others' strength of motivation to avoid acting inconsistently with this value. Thus they conclude that others' norm-following actions accurately represent their private preferences, though in fact those actions may be dispreferred (resulting instead from the individuals' priors and observations of others' behavior). One example of this effect is the reluctance of schoolchildren to risk 'looking stupid' by asking questions in a classroom situation (Miller and McFarland 1987): they assume that others do not ask questions because they understand the material, when in fact it is simply to avoid potential embarrassment. Thus the children keep quiet, not realizing that their silence results from, and adds to, a cascade effect which influences others to keep quiet as well. In other words, pluralistic ignorance causes individuals to assume that their observations are independent, though in fact they may be strongly correlated due to the effects of cascades. Bicchieri and Fukui go on to argue that inefficient norms result from cascade effects under conditions of pluralistic ignorance, and present a model of this norm-following behavior. In fact, the assumption of pluralistic ignorance is not absolutely necessary for cascades to occur, though it does tend to strengthen their effects. If pluralistic ignorance does *not* hold, individuals (realizing that cascade effects occur) may place less weight on their observations, reducing the probability and magnitude of cascades. Nevertheless, the assumption of pluralistic ignorance is very reasonable in our model. We assume a heterogeneous population in which individuals vary in their weight of public opinion; since the makeup of the population is *not* common knowledge, individuals cannot tell whether their observations resulted from a 'norm-ignoring' or 'norm-following' population. Thus the assumption of pluralistic ignorance is simply equivalent to assuming that individuals believe a priori that z (the proportion of the population dependent on public opinion) is low, and this is a perfectly reasonable thing for them to believe.

A third reason why agents may not take cascade effects into account is that, in certain cases, the realization that cascades occur may not affect an agent's decision. If the agent's weight of public opinion is sufficiently high, he will follow the decision that is more likely to be in the majority, regardless of the possibility that his observations may be wrong, and regardless of the size of the majority. Though cascades may cause overestimates of the size of the majority, and may occasionally lead to incorrect beliefs about the majority preference, the perceived majority preference is still *more likely* to be the true majority preference, and thus will be chosen by individuals with high weight of public opinion, even if pluralistic ignorance does not hold. For example, consider the basic informational cascade model of Bikhchandani et al. (1992), where every individual has partial information about the 'correct' decision, and every individual's private information has equal precision. Thus, if the first two individuals choose an action a, every succeeding individual will also choose that action. This does not require pluralistic ignorance to occur: even if a rational individual *knows* that all decisions after the first two are the result of an informational cascade, it will still be in his best interest to follow their lead, since the information content of the first two decisions outweighs his own private information. Similarly, in our model we often make the assumption that every individual has either very high or very low weights of public opinion and observation: in this simplified case, an individual with high weights will follow the herd, and an individual with low weights will follow his private preference, regardless of whether cascades are taken into account.

A fourth reason why agents may not consider cascade effects is that their goal may actually be to match other agents' *actions* rather than their preferences. This may be the case when the main reason for agreeing with others is purely for coordination purposes rather than the result of social pressure to conform. For example, an individual would prefer to drive on the right side of the road if almost everyone drives on the right, even if most people would have preferred (in the absence of other traffic) to drive on the left: in this case it is clear that agreement with actions, not with preferences, maximizes an agent's utility. If an agent's utility is increased when his action agrees with other agents' actions, then our naïve decision rule corresponds to one of two assumptions: either a) agents only need to coordinate with previously made actions (i.e. future decisions are irrelevant to their payoffs, as in Arthur's (1989) model), or b) agents assume that past observations are an accurate predictor of future decisions. This latter assumption is somewhat reasonable since the probabilities of positive and negative cascades are dependent on the current value of ΔN; on the other hand, agents would then be *myopic* in the sense that they do not consider the effects of their own action on the future actions of others. In fact, in early stages of the sequential decision-making process, a single decision may greatly affect the probabilities of positive and negative cascades, and naïve agents fail to take this into account.

Thus we have considered several reasons why agents might not adjust their weight of observation to take into account the possibility of cascade effects. Since the structure of the game, and the decision rules of the other players, are not assumed to be common knowledge, agents might not have sufficient information to take cascades into account. Alternatively, it might not be necessary to take cascades into account: this may be the case when agents attempt to coordinate with the actions of the other players rather than their preferences, or when certain restrictions are placed on the type of player. Finally, the phenomenon of pluralistic ignorance may cause agents to reason naïvely. If none of these reasons are applicable to a given situation, and agents have unlimited rationality, it might be necessary to extend the micromodel to account for agents' belief that cascades may occur. We prefer to assume that at least one of the above reasons is valid, and thus that agents make rational, but myopic, decisions as given by our model of 'naïve Bayesian norm followers.'

6. Conclusions

Thus we have presented a model of naïve Bayesian norm followers, rational agents whose subjective expected utility is increased by adherence to an established behavioral norm. A naïve Bayesian norm follower's estimate of the utility of an action is increased proportional to the probability that the action will agree with that of another (randomly selected) agent, and this probability is calculated using Bayes' rule. However, agents are 'naïve' in that they assume independence of previously made decisions; we consider a variety of phenomena, including pluralistic ignorance and lack of common knowledge, which may lead to agents' naïveté. We consider a heterogeneous population where different agents may be influenced to different extents by their private preferences, prior beliefs, and empirical observations of other agents' preferences; this generalizes previous models such as Banerjee (1992) and Bikhchandani et al. (1992), which assume a high degree of conformity (i.e. that every agent will follow a norm once it is established), equal priors, and common knowledge of priors. Our model can be applied both to strong cascades (where agents' private preferences are weighted proportional to the preferences of q_0 other agents) and weak cascades (where agents' private preferences are weighted proportional to the preference distribution of society as a whole), and is applicable to cascade effects resulting from a variety of positive feedback mechanisms, including informational, reputational, and networking cascades.

In addition to presenting a model of how rational norm-following decisions can be made, we also investigated when unpopular and inefficient norms might emerge from sequential choice in a population of naïve Bayesian norm followers. Using a biased random walk model, we demonstrated that unpopular norms can result from 'negative cascades,' in which the majority of the population chooses a dispreferred action because each agent's observations lead him or her to incorrectly believe that he or she is following the majority preference. We have shown that unpopular norms can emerge even when the majority of agents do not place high weight on popular opinion, and under a wide range of conditions, including heterogeneity in preferences, priors, and the weights of public opinion and observation. We examined the effects of real-world phenomena such as marketing, fanaticism, and pluralistic ignorance on the model, and investigated how these factors affect the probability and magnitude of negative cascades. For example, marketing may influence the priors of individuals, convincing them that a particular product is majority-preferred; we considered the impact of both 'mass marketing' and 'viral marketing' techniques on consumer choice. We also examined various generalizations of the model, including preference-dependent parameters, simultaneity in choices, and varying weights of public opinion and observation. This allows us to present a simple but general model of norm-following behavior which explains the emergence and persistence of unpopular, inefficient behavioral norms in society.[9]

NOTES

1. Granovetter and Soong (1986) and others have also examined *dispersion effects*, where individuals are motivated by negative feedback to choose differently from previous decision-makers (for example, avoiding a restaurant because it is too crowded). Such effects are outside the scope of this paper.
2. To be more precise, some informational effects may be at work even in a subjective choice setting, but these are unlikely to be as influential as reputational effects. In our example of soft drink choice, an informational effect can occur if individuals are influenced by others' choices to conclude that a soft drink is better tasting or otherwise superior. In this case, others' choices are *perceived* as carrying information about product quality, even if no such information is actually present. This effect is closely related to the availability cascade model of Kuran and Sunstein (1999): just as the availability of a perception in public discourse may lead to belief in its plausibility, so might popularity of a product lead to belief in its innate worth.
3. The beta distribution is a special case of the Dirichlet prior distribution; this very general prior is commonly used for Bayesian estimates of multinomial probabilities from data when no other constraints are known. This method is also called 'maximum a-posteriori' (MAP) learning, or Laplace smoothing, in the statistical learning literature. Human behavior consistent with Bayes' rule has been observed

in laboratory experiments by Anderson and Holt (1997), but also see Kahneman and Tversky (1973, 1974) for situations where non-Bayesian behavior may occur.

4. An identical expression for $P(A)$ can be derived using *model averaging* rather than *model selection*, i.e. setting $P(A)$ and $P_0(A)$ equal to the expectations rather than the modes of their respective beta distributions. In this case, we have $P(A) = E_\theta[\Pr(\theta|N_A, N_B)] = (N_A + \alpha)/(N + \alpha + \beta)$. Then setting $P_0(A) = (\alpha)/(\alpha + \beta)$ and $k = (1)/(\alpha + \beta)$, we again obtain $P(A) = (P_0(A) + kN_A)/(1 + kN)$.
5. Arthur and Lane (1993) present an alternative Bayesian model in an informational cascade-type framework with a homogeneous population and various other strong assumptions (for instance, that individuals reveal not only their product choice but their estimate of the true value of that product). Even in this less general model, naïveté is assumed, as the agents do not model the market-share allocation process (how the sequence of choices depends on the products' values) but instead use a uniform prior. In other Bayesian models of informational cascades, such as Banerjee (1992) and Banerjee and Fudenberg (2004), agents are not naïve in this sense, but instead the models rely on extremely strong assumptions of common knowledge (e.g. common knowledge of rationality, common priors, and common knowledge of the Bayesian update rules) which limit their applicability outside a very specific informational cascade setting.
6. This assumption of a sequential, exogenously determined order of decisions is identical to the models of Arthur (1989), Banerjee (1992) and Bikhchandani et al. (1992). Other models, such as Chamley and Gale (1994), Caplin and Leahy (1994), Zhang (1997), and Bicchieri and Fukui (1999) allow agents to choose the timing of their decisions: as a result, agents with more significant information, or stronger preferences, tend to move first, and other individuals follow their lead.
7. In addition to influencing agents' prior probabilities, advertising also impacts consumer choice in several other ways. It may influence agents' private preferences (for example, by convincing them that the product is of superior quality), and may introduce agents to choices of which they were previously unaware. We neglect these two effects for the purposes of our discussion.
8. Incorrect information may still be obtained in the simultaneous choice situation due to sampling error. Even if the sample is unbiased, sampling results in significant variance for small sample sizes. In fact, in the 'trendsetters and conformists' model of Bicchieri and Fukui (1999), inefficient norms result from sampling variance rather than from cascade effects in the standard sense: the game is structured so that all trendsetters move first (simultaneously), choosing their private preference, then all conformists move (simultaneously), following the norm set by the majority of trendsetters. A negative cascade can result when the trendsetters are not a representative sample of the population, i.e. when the majority preference of the trendsetters differs from the majority preference of the population as a whole. In the absence of a systematic bias in preferences, the probability of negative cascades approaches zero when the number of trendsetters becomes large; thus the assumption that the number of trendsetters is small is crucial to the Bicchieri and Fukui model. Our model, on the other hand, does not rely on a small, predetermined set of trendsetters: instead, negative cascades may emerge spontaneously from the aggregation of information inherent in sequential choice situations. We believe that this is a more reasonable model of the cascade effects observed in voting, product choice, and other situations where negative cascades may lead to inefficient and unpopular norms.

9. The author wishes to express his thanks to Cristina Bicchieri, Andrew Moore, Weng-Keen Wong, and three anonymous referees for their insightful and helpful comments.

REFERENCES

Anderson, L. R. and Holt, C. A. 1997. 'Information Cascades in the Laboratory.' *American Economic Review* 87: 847–62.

Arthur, W. B. 1989. 'Competing Technologies, Increasing Returns, and Lock-in by Historical Events.' *The Economic Journal* 99(394): 116–31.

Arthur, W.B . and Lane, D. 1993. 'Information Contagion.' *Structural Change and Economic Dynamics* 4: 81–104.

Banerjee, A. V. 1992. 'A Simple Model of Herd Behavior.' *Quarterly Journal of Economics* 107: 797–817.

Banerjee, A. V. and Fudenberg, D. 2004. 'Word of Mouth Learning.' *Games and Economic Behavior* 46: 1–22.

Bicchieri, C. 2003. *The Grammar of Society*. Cambridge, UK: Cambridge University Press.

Bicchieri, C. and Fukui, C. 1999. 'The Great Illusion: Ignorance, Informational Cascades, and the Persistence of Unpopular Norms.' *Business Ethics Quarterly* 9: 127–55.

Bicchieri, C. and Rovelli, C. 1995. 'Evolution and Revolution: The Dynamics of Corruption.' *Rationality and Society* 7: 201–24.

Bikhchandani, S. et al. 1992. 'A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades.' *Journal of Political Economy* 100: 992–1026.

Caplin, A. and Leahy, J. 1994. 'Business as Usual, Market Crashes, and Wisdom after the Fact.' *American Economic Review* 84: 548–65.

Chamley, C. and Gale, D. 1994. 'Information Revelation and Strategic Delay in a Model of Investment.' *Econometrica* 62: 1065–85.

Dawid, P., 1985. 'Clio and the Economics of QWERTY.' *American Economic Review Proceedings* 75: 332–7.

Domingos, P. and Richardson, M. 2001. 'Mining the Network Value of Customers.' *Proc. 7th Intl. Conference on Knowledge Discovery and Data Mining*: 57–66.

Granovetter, M. 1978. 'Threshold Models of Collective Behavior.' *American Journal of Sociology 83: 1420-1443*.

Granovetter, M. and Soong, R. 1983. 'Threshold Models of Diffusion and Collective Behavior.' *Journal of Mathematical Sociology* 9: 165–79.

Granovetter, M. and Soong, R. 1986. 'Threshold Models of Interpersonal Effects in Consumer Demand.' *Journal of Economic Behavior and Organization* 7: 83–99.

Granovetter, M. and Soong, R. 1988. 'Threshold Models of Diversity: Chinese Restaurants, Residential Segregation, and the Spiral of Silence.' *Sociological Methodology* 18: 69–104.

Kahneman, D. and Tversky, A. 1973. 'On the Psychology of Prediction.' *Psychological Review* 80: 237–51.

Katz, M. and Shapiro, C. 1986. 'Technology Adoption in the Presence of Network Externalities.' *Journal of Political Economy* 94: 822–41.

Kuran, T. 1995. *Private Truths, Public Lies*. Cambridge, MA: Harvard University Press.

Kuran, T. and Sunstein, C.R. 1999. 'Availability Cascades and Risk Regulation.' *Stanford Law Review* 51: 683–768.

Lee, I. H. 1998. 'Market Crashes and Informational Avalanches.' *Review of Economic Studies* 65: 741–59.

Leibenstein, H. 1950. 'Bandwagon, Snob, and Verblen Effects in the Theory of Consumers' Demand.' *Quarterly Journal of Economics* 64: 183–207.

Lohmann, S. 1994. 'The Dynamics of Informational Cascades: The Monday Demonstrations in Leipzig, East Germany, 1989–91.' *World Politics* 47: 42–101.

Miller, D. T. and McFarland, C. 1987. 'Pluralistic Ignorance: When Similarity is Treated as Dissimilarity.' *Journal of Personality and Social Psychology* 53: 298–305.

Mitchell, T. 1997. *Machine Learning*. New York: McGraw-Hill.

O'Gorman, H. J. 1975. 'Pluralistic Ignorance and White Estimates of White Support for Racial Segregation.' *Public Opinion Quarterly* 39: 313–30.

Prentice, D. A. and Miller, D. T. 1993. 'Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm.' *Journal of Personality and Social Psychology* 64: 243–56.

Scharfstein, D. S. and Stein, J. C. 1990. 'Herd Behavior and Investment.' *American Economic Review* 80: 465–79.

Shiryayev, A. N. 1984. *Probability*. New York: Springer-Verlag.

Tversky, A. and Kahneman, D. 1974. 'Judgment Under Uncertainty: Heuristics and Biases.' *Science* 185: 1124–31.

von Neumann, J. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Watts, D. J. 2002. 'A Simple Model of Global Cascades on Random Networks.' *Proceedings of the National Academy of Sciences* 99: 5766–71.

Welch, I. 1992. 'Sequential Sales, Learning, and Cascades.' *Journal of Finance* 47: 695–732.

Zhang, J. 1997. 'Strategic Delay and the Onset of Investment Cascades.' *RAND Journal of Economics* 28: 188–205.

DANIEL B. NEILL is a Ph.D. student in the Department of Computer Science at Carnegie Mellon University. His research interests include game theory, artificial intelligence, machine learning, and data mining.

ADDRESS: Department of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891, USA [email: neill@cs.cmu.edu].