Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades

Sushil Bikhchandani; David Hirshleifer; Ivo Welch

Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades

Sushil Bikhchandani, David Hirshleifer and Ivo Welch

I n 1995, management gurus Michael Treacy and Fred Wiersema secretly purchased 50,000 copies of their business strategy book *The Discipline of Market Leaders* from stores across the nation. The stores they purchased from just happened to be the ones whose sales are monitored to select books for the *New York Times* bestseller list. Despite mediocre reviews, their book made the bestseller list. Subsequently, the book sold well enough to continue as a bestseller without further demand intervention by the authors.[1] Presumably, being on a bestseller list helps a book sell more because consumers and reviewers learn from the actions of previous buyers.

Reports of the actions or endorsements of one set of economic decisionmakers often influence the reactions and purchases of others. The transformation of New York's Times Square after long decay was triggered by an investment by Disney, after which "wait-and-see investors piled in" ("A Star is Reborn," 1996). Often there are opportunities to manipulate the process by which individuals learn from their predecessors. There is a word, "claque," for those hired to applaud loudly

---

[1] It is difficult to calculate the net returns on Treacy and Wiersema's investment, partly because it is likely that they were able to limit the costs by returning books to the publisher, and partly because the bestseller status of the book helped them to obtain speaking and consulting income ("Did Dirty Tricks Create a Bestseller?" 1996).

■ *Sushil Bikhchandani is Professor of Decision Sciences and Ivo Welch is Professor of Finance, Anderson Graduate School of Management, University of California, Los Angeles, California. David Hirshleifer is the Merwin H. Waterman Professor of Finance, University of Michigan Business School, Ann Arbor, Michigan.*

(or to heckle competitors) at musical and stage performances. Ancient Roman families hired professional mourners at funerals. Hennessy Cognac hired actors and models to order their product at fashionable bistros ("The New Hucksterism," 1996). Many of us identify restaurant quality by the fraction of seats occupied; perhaps not coincidentally, restaurants often close off back-room peak-load seating capacity until the main and most visible section becomes quite full. Advertisements report the fractions of doctors or dentists that use certain medications and health products.

We will argue in this essay that learning by observing the past decisions of others can help explain some otherwise puzzling phenomena about human behavior. For example, why do people tend to converge on similar behavior, in what is known as "herding"? Why is mass behavior prone to error and fads? We will further argue that the theory of observational learning has much to offer economics and business strategy.

Social observers have long recognized imitation as important in human society. Machiavelli (1514) wrote: "Men nearly always follow the tracks made by others and proceed in their affairs by imitation." The philosopher Eric Hoffer (1955) asserted: "When people are free to do as they please, they usually imitate each other. . . . A society which gives unlimited freedom to the individual, more often than not attains a disconcerting sameness."

This predisposition to imitate is deeply rooted. Gibson and Hoglund (1992) describe evidence that animals imitate each other in choices of mate and territories; for example, female guppies are more likely to choose males to mate with whom they have observed being selected by previous females. The propensity to imitate is presumably an evolutionary adaptation that has promoted survival over thousands of generations by allowing individuals to take advantage of the hard-won information of others. Within minutes of birth, human infants mimic the observed facial expressions of adults. As we grow older, we continue to be influenced by the observed actions of others, from the acquisition of Beanie babies and consumption of Prozac, to wider lifestyle, work, and recreation choices.

The simplest and most basic cause of convergent behavior is that individuals face similar decision problems, by which we mean that people have similar information, face similar action alternatives, and face similar payoffs. As a result, they make similar choices. If Ford simply makes a better car than Yugo, and consumers understand this, then they end up buying the same car. Of course, opposing tastes can lead to opposing actions even if information is similar; vegetarians and meat-lovers frequent different restaurants.

Herding may arise when payoffs are similar even if initial information is not. In this case people communicate with each other or observe the actions of others— or the consequences of these actions. The key issue is how individuals determine which alternative is better. Each individual could decide by direct analysis of the alternatives. However, this can be costly and time-consuming, so a plausible alternative is to rely on the information of others. Such influence may take the form of

direct communication and discussion with, or observation of others. We will call influence resulting from rational processing of information gained by observing others *observational learning* or *social learning*. This essay focuses mainly on the case where individuals learn by observing the actions of others.

There are several other possible causes of conformity which do not require great similarity in individuals' decision problems. These include positive payoff externalities, which lead to conventions such as driving on the right-hand side of the road; preference interactions, as with everyone desiring to wear the more "fashionable" clothing as determined by what others are wearing; and sanctions upon deviants, as with a dictator punishing opposition behavior.

A Model of Observational Learning

Observable Actions versus Observable Signals

We consider two scenarios. In both, each individual starts with some private information, obtains some information from predecessors, and then decides on a particular action. In the *observable actions* scenario, individuals can observe the actions but not the signals of their predecessors. We compare this to a benchmark *observable signals* scenario in which individuals can observe both the actions and signals of predecessors.[2]

Consider an example in which risk neutral individuals decide in sequence whether to adopt or to reject a possible action. The payoff to adopting, V, is either 1 or −1 with equal probability; the payoff to rejecting is 0. In the absence of further information, both alternatives are equally desirable. The order in which individuals decide is given and known to all.

Each individual's signal is either *High* or *Low*, and High is relatively more likely when adoption is desirable ($V = 1$) than when it is undesirable ($V = -1$). Specifically, each individual observes High with probability $p > 1/2$ if $V = 1$, and with probability $1 - p$ if $V = -1$. To put it a little differently, after observing only one High, an individual's posterior probability that $V = 1$ is p, and the probability that $V = 1$ is only $1 - p$ if he observes Low. (This can be confirmed with a calculation using Bayes' rule.) Thus, p is the (posterior) probability that the signal is correct. All private signals are identically distributed and independent conditional on V. Naturally, an individual's posterior belief about V also depends on information derived from predecessors, in ways that will differ in the two scenarios.

In the observable-signals scenario, the information signals enter the pool of public information one at a time as individuals arrive. Because all past signals are publicly observed, information keeps accumulating so that individuals, all of whom have the same payoffs from taking the same action, eventually settle on the correct

---

[2] See Welch (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Banerjee (1992).

choice and thus behave alike. If others' signals are observed with some noise, then information accumulates more slowly but still draws individuals toward the same, correct action.

Because actions reflect information, it is tempting to infer that if only the actions of predecessors are observable, the public information set will also gradually improve until the true value is revealed almost perfectly. However, we now show that a scenario of observable actions is actually quite different from a scenario of observable signals. In the observable-actions case, individuals often converge on the same wrong action—that is, the choice that yields a lower payoff. Furthermore, behavior is idiosyncratic, in the sense that the error-prone choices of a few early individuals determine the choices of all successors.

Returning to our example, the first individual, Aaron, adopts if his signal is High and rejects if it is Low. All successors can infer Aaron's signal perfectly from his decision: if he adopted then he must have observed High and if he rejected he must have observed Low. Now consider the choice of the second individual, Barbara. If Aaron adopted, then Barbara should also adopt if her private signal is High; as Barbara sees it, there have now been two high signals, the one she inferred from Aaron's actions and the one she observed privately. However, if Barbara's private signal is Low, then as she sees it, there has been a High signal (inferred from Aaron's actions) and her own Low signal, and she is exactly indifferent between adopting and rejecting. We assume, for expositional simplicity, that as Barbara is indifferent between the two alternatives, she tosses a coin to decide. (By similar reasoning, if Aaron rejected, then Barbara should reject if she observes Low, and toss a coin if her signal is High.)

The third individual, Clarence, faces one of three possible situations: both predecessors adopted, both rejected or one adopted and the other rejected. In the first case, where both his predecessors adopted, Clarence also adopts. He knows that Aaron observed High and that it is more than likely that Barbara observed High too (although she may have seen Low and flipped a coin). Thus, even if Clarence sees a Low signal, he adopts, because he believes that there is better than an even chance that the value to adoption is 1.[3] Consequently, Clarence's decision to adopt provides no information to his successors about the desirability of adopting. The fourth individual, Donna, finds herself in a similar situation as Clarence and adopts regardless of her signal, as will all her successors. Clarence is said to be in an *informational cascade*, because his optimal action does not depend on his private information, and the uninformativeness of Clarence's action means that no further information accumulates. Everyone after Clarence faces the same decision and also adopts based only on the observed actions of Aaron and Barbara. We therefore call this situation an Up cascade. Similarly, in the case where both Aaron

---

[3] If Clarence takes into account only Aaron's High signal and his own Low signal then he believes that the value to adoption is equally likely to be 1 or −1. But Clarence also knows that Barbara is more likely to have seen a High signal than a Low signal. This tilts the decision in favor of adoption.

and Barbara had rejected, Clarence and all successors reject even if they all privately observed High signals. This is a Down cascade.

In the remaining case where Aaron adopted and Barbara rejected (or vice versa), Clarence knows that Aaron observed High and Barbara observed Low (or vice versa). Thus, Clarence's belief based on the actions of the first two individuals is that the High and Low outcomes are equally likely. He finds himself in a situation similar to that of Aaron, so Clarence's decision is based only on his private signal. Then, the decision problem of Donna, the next in line, is the same as Barbara's. Aaron's and Barbara's actions have offset and thus carry no information to the fifth individual, Edgar. And if Clarence and Donna both take the same action—say, adopt—then an Up cascade starts with Edgar.

An individual's optimal decision rule may be summarized as follows. Let d be the difference between the number of predecessors who adopted and the number who rejected. If $d > 1$, then adopt regardless of private signal. If $d = 1$, then adopt if private signal is High and toss a coin if signal is Low. If $d = 0$, then follow private signal. The decisions for $d = -1$ and $d < -1$ are symmetric. The net preponderance of adoptions over rejections evolves randomly, and sooner or later, usually quite quickly, must bump into the upper barrier of $+2$ and trigger an Up cascade, or the lower barrier of -2 and trigger a Down cascade. With virtual certainty, all but the first few individuals end up doing the same thing.

Order of Information, Noise, and Information Externalities

The fundamental reason the outcome with observable actions is so different from the observable-signals benchmark is that once a cascade starts, public information stops accumulating. An early preponderance towards adoption or rejection causes subsequent individuals to ignore their private signals, which thus never join the public pool of knowledge. Nor does the public pool of knowledge have to be very informative to cause individuals to disregard their private signals. As soon as the public pool becomes even modestly more informative than the signal of a single individual, the next individual defers to the actions of predecessors and a cascade begins.

Furthermore, the type of cascade depends not just on how many High and Low signals arrive, but the order in which they arrive. For example, if signals arrive in the order HHLL . . . , then all individuals adopt, because Clarence begins an Up cascade. If, instead, the same set of signals arrive in the order LLHH . . . , all individuals reject, because Clarence begins a Down cascade. And if the signals arrive as HLLH . . . , then with probability one-half Barbara adopts and Clarence begins an Up cascade. Thus, in the observable-actions scenario, whether individuals on the whole adopt or reject is path-dependent.

To see how likely it is that a cascade occurs, consider the situation in which private signals are very noisy; specifically, the probability that the signal is correct is $p = 0.51$. Then, there is approximately a 75 percent chance that an Up or Down cascade forms after the first two individuals! To see this, first suppose that

$V = 1$. An Up cascade occurs either when Aaron and Barbara both receive High (with probability $0.51 \times 0.51 = 0.2601$) or when Aaron receives High and Barbara receives Low, flips a coin, and chooses to adopt ($0.51 \times 0.49 \times 0.5 = 0.12495$). A Down cascade occurs either when Aaron and Barbara both receive Low (with probability $0.49 \times 0.49 = 0.2401$) or when Aaron receives Low, Barbara receives High, but flips a coin and decides to reject ($0.49 \times 0.51 \times 0.5 = 0.12495$). Summing these probabilities, a cascade occurs with slightly more than 75 percent of the time after the first two players. (A symmetric calculation applies if $V = -1$). Remember that if the actions of the first two players differ, then their information offsets so that the game effectively begins afresh with the third player; if the actions of the third and fourth players differ, then the game effectively begins afresh with the fifth player. After eight players the probability is only 0.004 that such offsetting has occurred four times, leaving a 0.996 probability that individuals are in a cascade.

When $V = 1$, the probability of an Up cascade after two individuals, based on summing the probabilities above, is 0.38505 (that is, $0.2601 + 0.12495$) while the probability of a Down cascade is .36505 (that is, $0.2401 + 0.12495$). So given that a cascade has occurred, the chance of it being a correct Up cascade rather than an incorrect Down cascade is 51.3 percent ($0.38505/[0.38505 + 0.36505]$). By similar reasoning, the probability of a cascade that started at any stage being correct is 51.3%. Compare this with a scenario in which individuals do not observe their predecessors at all. Then each individual would choose the right action, based only on the private signal, with a probability of 51 percent. In this case, the gain in accuracy from observing the actions of predecessors is a minimal 0.3 percentage points. In the observable-signals scenario, publicly observed information signals of predecessors are virtually conclusive as to the right action after many individuals. In contrast, when only actions are observed, decisions are little better than when individuals cannot observe predecessors at all.

More generally, even when individuals have more accurate signals, the information contained in a cascade is not substantially better than a single individual's signal. Figure 1 illustrates the point. The horizontal axis shows p, the probability that the signal is correct. In the long run a cascade will eventually occur, which will be either correct or incorrect; the vertical axis shows the probabilities that a correct cascade or that an incorrect cascade eventually occurs. Thus, when p = 0.7, the probability of an eventual correct cascade is 0.753, and the probability of an eventual incorrect cascade is the remaining .247; for p = 0.8, the probability of an eventual correct cascade is 0.857, and the probability of an eventual incorrect cascade is .143.

When an individual takes an action that is informative to others, it provides a positive externality. This desirable information externality is weaker when only past actions are observed than when past signals are observed, and once a cascade starts, the information externality disappears altogether. If an individual was expected to make the error of following the private signal instead of obeying

Figure 1
Cascade Probabilities
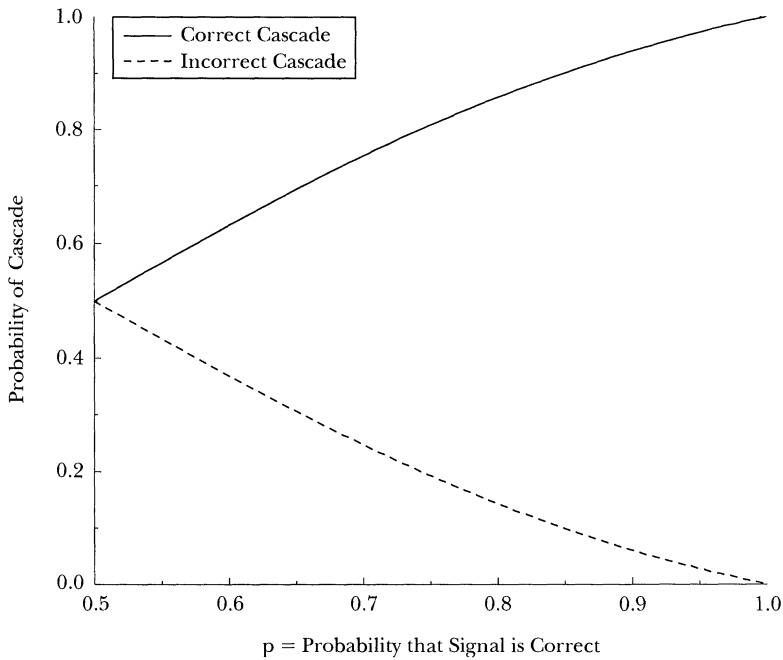
the cascade, the actions of that individual would add to the public pool of knowledge, to the benefit of followers. Such altruistic behavior by a number of individuals would ultimately lead to almost perfectly accurate decisions in the long run. Instead, individuals, acting in their own self-interest, rationally take uninformative imitative actions. Bernardo and Welch (1997) point out that irrationally overconfident entrepreneurs, who place heavy weight on their own signals relative to those of others, may be exceptionally useful citizens. More generally, the theory of informational cascades suggests that social misfits of various sorts—such as newcomers who have not observed past history, or prophets with special information sources—may disproportionately benefit society (Hirshleifer and Noah, 1997).

Fragility

Of course, in reality we do not expect a cascade to last forever. Several possible kinds of shocks could dislodge a cascade; for example, the arrival of better informed individuals, the release of new public information, and shifts in the underlying value of adoption versus rejection. Indeed, when participants know that they are in a cascade, they also know that the cascade is based on little information relative to

the information of private individuals. Thus, a key prediction of the theory is that behavior in cascades is *fragile* with respect to small shocks.[4]

To illustrate fragility, consider a modification of the basic example in which each individual usually receives one High or Low signal, or with a small probability, say 0.001, instead receives *two* conditionally independent draws of the signal. It is very likely that each of the first few individuals receives only one draw of the signal, and that a cascade starts. Suppose that this is an Up cascade. Ultimately, a one-in-a-thousand individual, whom we will call Spock, observes two signal draws. If Spock sees two Low signals, that is sufficient for him to go against the cascade and reject. This is because Spock knows something about four signals: the first one, which must certainly have been High; the second, which could have been High (though there is some chance that the second decisionmaker received a Low signal but flipped a coin and adopted anyway); and Spock's own two draws, both Low. All of the intervening actions from the third individual up to Spock's predecessor were part of the cascade, and thus their actions revealed no information. Based on the two Low signals, choosing to reject is logical for Spock. This dislodges the cascade, as successors correctly infer that Spock observed two Low signals.

Recall that if $p = 0.51$ there is a 0.487 chance that the original Up cascade was incorrect. In this case, the unconditional probability that Spock observes two Low signals and overturns the Up cascade is a high 0.24984. A new cascade develops soon thereafter. If the next person draws a Low signal, then a Down cascade is started. But if the next person draws a High signal, then it may take several more draws before a cascade reasserts itself. This new cascade may again be overturned later by an individual who receives two signals.

So far we have argued that cascades are born quickly and idiosyncratically, and shatter easily. How robust are these conclusions? When some assumptions in the example are relaxed, is the aggregation of information still inefficient or delayed?

Informativeness of Past Actions

Often only a summary statistic of the actions of predecessors is observable. For example, an individual may learn that the prescription medicine Tagamet is outselling Pepcid, without knowing the order in which individuals purchased. (In fact, a SmithKline Beecham advertising campaign in 1985 stated that their product Tagamet had racked up 237 million prescriptions versus Pepcid's 36 million.) The observability of summary statistics still leads to idiosyncratic outcomes, fragility, and cascades. The basic intuition is as before. Information keeps accumulating until a preponderance of evidence supports one action or the other by just enough to outweigh one individual's private signal. At this point a cascade starts and new information stops accumulating.

---

[4] There are some models enforced by the threat of sanctions upon defectors in which rare shifts occur when the system crosses a critical value that shifts the outcome from one equilibrium to another (Kuran, 1989).

A related situation occurs when individuals have the opportunity to observe only a few predecessors, such as neighbors, instead of the whole chain. For example, Rogers (1983) reports that agricultural innovations were influenced heavily by choices of neighbors. This leads to similar outcomes, as long as enough predecessors can be observed. (For instance, in the example above, observing only a single predecessor may not provide enough information to start a cascade, but observing two does.)

If there are more than just two possible action alternatives, informational cascades can still result. However, as the set of alternatives becomes larger and richer, cascades tend to take longer to form and aggregate more information. If the set of action alternatives is continuous (for example, all points on the interval [0, 1]), then even an individual late in the sequence will still adjust his action at least slightly based on the private signal (Lee, 1993). Consequently, private signals can be perfectly inferred from actions, information aggregates efficiently, and cascades do not form. However, if individuals cannot distinguish between nearby actions taken by their predecessors, cascades do arise.

This reasoning suggests that cascades are most important for phenomena that have an important element of discreteness or finiteness. For example, investment projects have a minimum efficient scale, leading to a discrete difference between not investing and investing. Votes are taken between a discrete set of alternatives. Consumers cannot choose a car halfway between a Ford and a Toyota, a potential acquirer bids or does not bid for a target firm, and an employee is either hired or fired.

Furthermore, when individuals have bounded powers to perceive or recall fine gradations, they tend to divide up actions into discrete choices, even when those actions have a continuous character. Verbal concepts combine separate items into coarser categories. We remember a color as "red" rather than the exact shade of red. We think of people as honest or dishonest, distinguish friends from acquaintances and enemies, and for that matter think of statements as "true" or "false." The categorizing inherent in ordinary conversation suggests that cascades can form even when individuals can credibly communicate with each other verbally, because much information is transmitted as discrete categories.

Discreteness or finiteness can be viewed as a way of adding noise or distortion to past signals. The main contribution of the informational cascades theory is to show that when individuals see past signals only through a crude discrete filter—for example, whether an action was adopted or rejected—then learning is surprisingly imperfect and can quickly become completely blocked. Discreteness is of course not the only way to add noise to the observation of past signals; for example, there could instead be direct noise in observation of past actions (Vives, 1993; Cao and Hirshleifer, 1997a). Such noise slows down the rate of learning, but if actions are continuous, learning is not completely blocked. Still, either way, information aggregation is inefficient, wrong actions are sometimes taken for a long time, and the path to convergence is idiosyncratic.

In contrast with the requirement of discrete or finite actions, informational cascades do not require any discreteness in the information signals received by individuals. However, for cascades to arise, signals must not be conclusive. After all, if an individual receives a signal realization so informative that it provides virtually perfect information about the true value, the individual follows it without regard to the actions of predecessors. If such signals are always possible, individuals ultimately converge upon the correct action. However, if virtually conclusive signal values are rare, actions may be mistaken for a long time.

Differing Information Precision: Fashion Leaders

Up to this point, individuals have been assumed to be identical, except for the different signal draws they may receive. Of course, individuals actually differ in many dimensions, including their preferences, payoffs, and the precision of the information they receive. Allowing for such heterogeneity can either exaggerate or moderate cascading behavior.

Consider, for example, several neighbors deciding between a Ford and a Toyota. One is a car mechanic, and therefore better informed than the others about which alternative is better. If the mechanic chooses relatively late in the decision queue, he can break an existing cascade because he may follow his own signals rather than defer to predecessors. Suppose, however, that the first decisionmaker Aaron is the well-informed mechanic. In this case, Barbara immediately defers to Aaron's decision, and a cascade forms instantly—Aaron is a "fashion leader." Social psychologists report that people imitate the actions of those who appear to have expertise. This is probably part of what underlies the success of product endorsements in which athletes are seen to use a particular brand of athletic shoes or tennis racket.

This drawback of leading off with the best informed has not been lost on designers of judicial systems. According to the Talmud, judges in the ancient Hebrew Sanhedrin (high court) voted on cases in inverse order of seniority to reduce the natural influence of older (and presumably wiser) judges on the choices of junior judges. Similarly, in U.S. Navy courts-martial, judges vote in inverse order of rank.[5] In simultaneous balloting, voters decide without knowing how others have voted. Thus, the advantage of having committee members cast a simultaneous ballot instead of a public, sequential ballot is that it leads to more informed decisions.

Differing Preferences and Payoffs: To Each His Own

What if different individuals value adoption differently? Suppose that individuals are classified into two or more types according to their preferences; equivalently, imagine that the payoffs from adopting differ for each individual. As an

---

[5] An alternative explanation is that junior judges may think that conforming with their superiors is good for their careers. These institutions also reduce the incentives for such opportunistic behavior.

extreme case, consider opposing preferences or payoffs, where individuals prefer opposite behaviors. For example, a new age vegetarian may want to avoid the restaurant favored by the football team, and vice versa. If each individual's type is observable, then until cascades start, an individual's action together with the individual's type conveys information about the signal received by that individual. As with the case of homogeneous individuals, cascades start when information in the history of predecessors' actions outweighs an individual's private signal. Late-deciding individuals of the same type will eventually choose the same action regardless of their private information, but different types may cascade on different actions.

However, if the type of each individual is only privately known, and if preferences are downright opposing, then learning may be confounded because individuals do not know what to infer from the mix of preceding actions they observe (Smith and Sorenson, 1995). More typically, even when preferences and payoffs are not completely opposing, uncertainty about the characteristics of predecessors can slow the rate of learning. For example, a software writer may commit to the Java platform either because she is optimistic about its prospects (favorable signal realization), because she is relatively tolerant of risk or enjoys writing programs using this approach (heterogeneous preferences), because she thinks her firm's own profits will be particularly high if Java catches on (heterogeneous payoffs), or because she has made a mistake (imperfect rationality). A later individual can't be sure why she has adopted early. This makes the actions of early decisionmakers more noisy as indicators of their signals. Nevertheless, if enough writers adopt Java, the evidence implicit in their actions will convince even doubters with opposing signals. The bottom line is that, although it may take longer when actions are noisy, as long as individuals' action sets are not continuous and unbounded, cascades form when the public information set has become precise enough to outweigh an individual's private signal in determining his action.

Changing Tastes or Payoffs

Suppose that instead of a constant underlying value, there is a small probability that the payoff value may change each time period. Then cascades can still occur. However, since cascades aggregate very little information, at some later point in time large changes in behavior may occur without a readily apparent reason; these shifts in behavior are driven by an expectation that the payoff value has changed. Such seemingly whimsical shifts in behavior appear faddish. Furthermore, as Perktold (1996) shows, information aggregation remains inefficient.

Timing Choice and the Explosive Onset of Cascades

Sometimes large groups of people adopt new behaviors with startling rapidity. From teenage mutant Ninja turtles to oat bran fads, from counter-culture movements to religious revivals, the timing of such sudden changes is usually unpredictable. As with the example of wait-and-see real estate investors piling into New York's Times Square after Disney, giving people the choice of when to act can lead

to sudden onset of cascades wherein many followers simultaneously adopt a new behavior.

Suppose that at each instant all individuals who have not yet chosen an action may adopt, reject, or as a third choice, delay making a decision.[6] There is a small cost per unit time of postponing the decision. Individuals differ slightly in the reliability of their High or Low signals; that is, when $V = 1$ one individual observes High with a greater probability than another. Higher precision individuals (like the car mechanic discussed above) have less to gain from waiting to see the actions of informational inferiors, so they tend to move first. If signal accuracy is not public knowledge, then subsequent individuals can infer the accuracy of the first individual's signal from the delay before action. They disregard their own noisier signals and copy the first individual's decision immediately. Thus, all actions are deferred until one individual triggers an explosion of simultaneous cascading activity. And since the highest precision individual decides first, this can lead to even more extreme idiosyncrasy in which all actions are based only on a single individual's information.

Costly Information, Alternative Information Sources, and Network Externalities

In the basic example, individuals received private information free of charge. If, instead, individuals have to pay a fixed cost to obtain private signals, cascades may form instantly, because Barbara may find it optimal to rely on Aaron rather than incur the investigation cost. Paradoxically, the ability to learn by observing predecessors can make the decisions of followers noisier by reducing their incentives to collect (perhaps more accurate) information themselves (Cao and Hirshleifer, 1997b).

Individuals often learn more than just past actions. It might be supposed that additional sources of information would tend to improve information aggregation and, perhaps, prevent cascades. After all, perfect observation of past signals would, of course, lead to socially (as well as privately) optimal choices. Indeed, there are circumstances where the ability to observe a random sample of past actions *and outcomes* leads to convergence to correct choices (Banerjee and Fudenberg, 1995). However, even when individuals can observe all past actions and resulting payoff outcomes, idiosyncratic cascades can still form (Cao and Hirshleifer, 1997b). For example, a string of early individuals may cascade upon alternative A, and its payoff may become visible to all, yet alternative B (whose payoff is still hidden) may be superior. Indeed, the ability to observe past payoffs can sometimes trigger cascades even more quickly.

We have assumed that individuals care about others' actions only because they convey information about the value of adoption. In many realistic settings, in ad-

---

[6] The following discussion is based on Hendricks and Koveneck (1989), Chamley and Gale (1994), Caplin and Leahy (1994), Gul and Lundholm (1995), and Zhang (1997).

dition to the informational externality described here, there are direct payoff interactions in the form of (positive) consumption or production externalities—sometimes called network externalities. The intuition here is that joining a network may help both the joiner and others who have already joined. Uniformity is likely in the presence of positive network externalities.[7] However, this uniformity does not display the fragility of an informational cascade. When there are positive network externalities and imperfect information about payoffs, observational learning can be pivotal early in the process in determining which behavior is fixed and reinforces the path-dependence of the outcome (Choi, 1997). Khanna (1998) examines how cascades affect managerial contracts and investment decisions among competing firms even without network effects.

Efficiency

We have shown that cascades weaken a favorable informational externality. Therefore outcomes are inefficient relative to the observable-signals scenario. This inefficiency arises from the discrete or bounded nature of possible actions, which limits information transmission. In principle, trade in information could solve these inefficiencies, but the transaction costs of buying information from scattered and unfamiliar predecessors could be quite high; further, there are problems of credibility, which lead to imperfect markets for information; after all, actions speak louder than words.

Potentially, a third party such as government could help by gathering and disseminating information. A less centralized approach would be to improve institutions and technologies by which individuals who face similar choices can identify each other and communicate their information. This consideration suggests that the rise of the internet (and intranets within organizations) will reduce the problems of cascades. However, there is an opposing effect: improved communication also help individuals learn about the actions of others. This may reduce an individual's incentives to gather information, allow cascades to start sooner and to extend to larger subsets of the entire decision-making population. Indeed, it would be socially most advantageous if one could isolate different groups of decisionmakers, and then disclose their actions simultaneously.

Applications

We now discuss some situations in which observational learning plays an important role.

---

[7] See the articles in the Symposium on Network Externalities in the Spring 1994 issue of this journal, and Arthur (1989).

Laboratory Experiments

Laboratory experiments provide the cleanest tests of social learning theories, since controls minimize potentially confounding affects. Anderson and Holt (1997) describe an experimental environment designed to test the basic cascades model. (Anderson and Holt (1996) describe how these experiments can be repeated in a classroom setting.) Subjects were rewarded for correctly guessing the urn from which a ball was drawn. All balls were drawn from the same urn (with replacement). One urn contained two-thirds black balls; the other, two-thirds white balls.

Each individual in a sequence observed the color of his ball, as well as the guesses of predecessors. In 94 cases, an individual was confronted with a situation in which it was optimal to follow the guess of that person's immediate predecessor in opposition to the person's own private signal. In other words, the individual was in an informational cascade. In 79 of these cases individuals acted against their own signals and followed the cascade.

Business Strategy

The theory of informational cascades theory suggests that firms should imitate each other in their product decisions. However, conventional industrial organization theory often implies that firms should differentiate their products to decrease competition and raise profit margins. Thus, observing uniform behavior in certain settings supports the hypothesis that observational learning is important.

Kennedy (1997) examines decisions by television networks to introduce different kinds of shows from 1960–89. The logic of product differentiation suggests that the introduction of a medical drama by ABC, for example, should reduce the benefit to NBC and CBS from doing so. However, if NBC and CBS believe that ABC has information about changing public tastes for different kinds of shows, they may want to imitate ABC's choice. After controlling for other factors, Kennedy finds that "the networks tend to make introductions in the same categories as their rivals (e.g., situation comedies, medical dramas, adventure series)." He concludes that "in at least one industry, strategic imitation appears to be common" contrary to "the more traditional differentiation hypothesis."

A potential problem with studies of imitation is that there can be common information signals—for example, about shifts in viewers' tastes—that are observable to the TV networks but not to the econometrician. This could lead to commonality of behavior without imitation. But Kennedy points out: "While of theoretical concern, conversations with programming analysts at both CBS and NBC indicate that no reliable common signal exists. Each network performs extensive market research, but there are no important independent sources of information (other than ratings, which are observed by the econometrician) and joint market research does not generally occur." Moreover, an obstacle to direct communication in this context is that networks are likely to be skeptical of any information offered to them by their competitors.

Is there a more general tendency toward strategic imitation? In Gilbert and

Lieberman's (1987) study of 24 chemical products over two decades, larger firms in an industry tend to invest when their rivals do not, but smaller firms "tend to follow the investment activity of others." This behavior is consistent with a "fashion leader" version of the cascades model in which the small free-ride informationally on the large.

In an example of spatial clustering of bank branches in cities, Chaudhuri, Chang and Jayaratne (1997) point out that banks may have imperfect information about the potential profitability of opening a branch in a particular neighborhood. They show that a bank's decision to open a new branch in a census tract of New York City during 1990–95 depended on the number of existing branches in that tract. They use tract-level socioeconomic data, land-use data, and crime statistics to control for expected tract profitability. Still, they report a positive incremental relation between a bank's decision to open a new branch and the presence of other banks' branches and conclude that the evidence supports information-based imitation.

There are instances in which hindsight suggests that incorrect cascades persisted for a time. Wooden plank toll roads originated in Russia and were introduced in Canada in 1840. In 1844, the promoter George Geddes convinced the town of Salina, New York, that plank roads would last about eight years. In 1846, the Salina road was completed, and 289 New York plank road companies incorporated in the following four years. Other promoters began to claim durability of even 10–15 years. Altogether, 10,000 miles of plank roads were constructed. The revelation of the true life-span of about four to five years came in 1852, when the Salina road deteriorated dangerously. Plank road construction quickly came to a halt (Klein and Majewski, 1996).

Consumer Marketing

We described earlier some questionable methods of manipulating social learning, such as inflating the sales measures used for constructing a bestseller list. The cascade theory explains why the ubiquitous and legitimate marketing method of offering a low initial price may be a successful scheme for introducing an experience good: early adoptions induced by the low price help start a positive cascade. This idea was first analyzed by Welch (1992) to explain why initial public offerings of equity are on average severely underpriced by issuing firms. Disney sells its movie videocassettes with special bonuses (in effect, price cuts) for advance buyers. Indeed, a seller may be tempted to cut price secretly for early buyers, so that later buyers will attribute the popularity of the product to high quality rather than low price.

Crime and Enforcement

There is a great deal of evidence that the decision to commit crime is influenced by observing the behavior of others; for an excellent discussion, see Kahan (1997). When individuals see peers commit crime, they may infer that others per-

ceive the probability of gain to be high and of punishment or stigmatization to be low. If apprehension is rare, a few individuals who are relatively insensitive to penalties and continue to commit visible crimes may lead to a broader inference in the community that crime pays.[8]

Evidence also suggests that the underlying determinants of crime are idiosyncratic. Sometimes public news of one kind of crime leads to more of that crime. Sheffrin and Triest (1992) report that news stories about tax non-compliance spark greater tax evasion by others. Several researchers have provided evidence of contagion in more spectacular crimes such as assassinations, hijackings, kidnappings, and serial murders (Bandura, 1973; Berkowitz, 1973; Landes, 1978). Other studies have found that crime is tied to whether others in the neighborhood are committing crime, even after controlling for demographic variables (such as race and income) and law enforcement. Glaeser, Sacerdote and Scheinkman (1996) provide evidence from New York City neighborhoods that individuals are more likely, ceteris paribus, to commit crimes when those around them do, controlling for a variety of variables. Skogan (1990) provide evidence from 40 urban neighborhoods that robbery rates are correlated with measures of social disorder (such as graffiti). In both studies, individuals' decisions to commit crime and the presence of gangs were more influenced by others than by demographic variables such as race and poverty, and by law enforcement. Several studies report that increased enforcement and penalties for gang crimes have been ineffective (Miller, 1990; Huff, 1990; Office of Juvenile Justice, 1994).

The social influence approach suggests alternative methods, such as curfews and anti-loitering laws that make gangs and criminality less visible. Obvious signs of crime (such as broken windows) influence perceptions about the likely consequences of more serious crime. Kahan (1997) argues that crime deterrence policies need to take such social influence into account and emphasizes the importance of "order maintenance." Ironically, conspicuous self-protection measures by private citizens, such as alarm systems or heavy bars and locks, can convey to others that criminality is rampant—and therefore presumably profitable. In 1993, the New York City Police Department began to enforce more aggressively the rules on public order offenses, such as vandalism, aggressive panhandling, public drunkenness, unlicensed vending, public urination and prostitution. Over the next three years, serious crimes in New York decreased sharply. The effectiveness of this strategy is puzzling under traditional theories of crime, less so under the social influence approach.[9]

Politics

People can learn about the political preferences of others by observing public protests, demonstrations, and even riots. According to Lohmann (1994), both the

---

[8] Moreover, the consequent increase in criminal activity may lead to an actual decrease in the probability of apprehension given the limited resources available for law enforcement. Hence, the perception that crime pays becomes a self-fulfilling prophecy.
[9] In a recent issue of this journal, DiIulio (1996) argues that existing theories of crime are inadequate and urges economists to come up with new alternatives.

threat of sanctions and informational cascades played a role in the maintenance and collapse of the East German regime. Secret opinion polls conducted by the Communist Party had shown widespread disapproval for years. However, the threat of arrest by state security police prevented individuals from publicly expressing their dissatisfaction, ensuring lack of protest. Information revelation eventually came about almost as an accident. The geography of Leipzig allowed people to congregate in a public plaza after church services. Weekend by weekend, the turnout of protesters in Leipzig ranged from 25 to 2500 per month in the first half of 1989, and exploded to 1.4 and 3.3 million in October and November of that year. At this point, the East German leader Erich Honecker publicly defended the Tienanmen Square actions of the Chinese government, and issued an "order to shoot." Large supplies of tear gas and special army troops were unloaded in Leipzig, and hospitals prepared for a bloodbath. However, Lohmann concludes that protesters inferred from the participation of others that the potential benefits (regime collapse) outweighed the costs (risk of a bloodbath).

People can also learn about others' political beliefs by observing polls and others' votes. This has led to the complaint that early reporting of election results or polls is undesirable, because early respondents carry disproportionate weight. Several European countries prohibit publication of poll results close to their election dates. Iowa voters gave an obscure candidate named Jimmy Carter a conspicuous early success in the 1976 U.S. presidential campaign. Many Southern states hold their primaries early in the election cycle on the same date ("Super Tuesday"), presumably to increase their influence on the presidential election.

Medical (Mal)practice

Most doctors cannot stay fully informed about relevant medical research advances in all areas. The theory of information cascades predicts fads, idiosyncrasy, and imitation in medical treatments. It has indeed been alleged that a blind reliance by physicians upon what colleagues have done or are doing commonly leads to surgical fads and even to treatment-caused illnesses (Robin, 1984; Taylor, 1979). Bleeding as a treatment, popular until the 19th century, is a familiar example. Many dubious practices seem to have been adopted initially based on weak information, such as elective hysterectomy (the routine surgical removal of the uterus of women past childbearing age), and tonsillectomy. Differences in tonsillectomy frequencies as well as other procedures in different countries and regions are extreme (Phelps and Mooney, 1993).

Concluding Remarks

There are many patterns of convergent behavior and fluctuations in the world that do not make immediate sense in terms of traditional economic models, such as fixation on wrong technologies, stock market crashes, sharp shifts

in investment and unemployment, bank runs, and reversals in election out-
comes. Such behavioral convergence often appears spontaneously without any
obvious punishment of defectors, sometimes even in the face of negative payoff
externalities.

Although other factors—such as network externalities and preference inter-
actions—can lock in an inefficient behavior, the informational cascades theory dif-
fers in that it implies pervasive but fragile herd behavior. This occurs because cas-
cades are triggered by a small amount of information. Under informational cas-
cades, the system spontaneously fluctuates until it reaches a precarious resting point
in which behavior is sensitive to small shocks.[10]

Most real applications involve mixtures of informational effects, sanctions
against defectors, network externalities, and preference effects. We believe that the
integration of learning/cascades effects with other factors will lead us to better
theories about the process by which society locks into technologies or customs, and
how information releases can be used to shift undesirable equilibria.

Observational learning theory suggests that in many situations, even if payoffs
are independent and people are rational, decisions tend to converge quickly but
tend to be idiosyncratic and fragile. Convergence arises locally or temporally upon
a behavior, and can suddenly shift into convergence on the opposite behavior. The
required assumptions, primarily discreteness or boundedness of possible action
choices, are mild and likely to be present in many realistic settings. This suggests
that cascade effects may be ubiquitous and have promise for explaining phenomena
that have puzzled economists and other social scientists.

---

[10] In this respect the cascades phenomenon is somewhat like physics models of "self-organized criticality"
(Bak and Chen, 1991). There are, however, some important differences. The most obvious is that the
basic elements of the cascades theory are rational, information-processing individuals. Also, there is a
broad parallel between cascades models and models of nonlinear dynamics (chaos theory) in that small
differences in initial conditions/realizations can make a large difference for later outcomes.

References

Anderson, Lisa, and Charles Holt, "Informa-
tion Cascades in the Laboratory," *American Eco-
nomic Review,* 1997, *87,* 847–62.

Anderson, Lisa, and Charles Holt, "Classroom
Games: Information Cascades," *Journal of Eco-
nomic Perspectives,* Fall 1996, *10*:4, 187–93.

Arthur, Bryan, "Competing Technologies, In-
creasing Returns, and Lock-in by Historical
Events," *The Economic Journal,* 1989, *99,* 116–31.

Bak, Per, and Chen, Kan, "Self-Organized
Criticality," *Scientific American,* January 1991,
264:1, 46–53.

Bandura, Albert, Aggression: A Social Learning Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Banerjee, Abhijit, "A Simple Model of Herd Behavior," *Quarterly Journal of Economics,* 1992, *107,* 797–818.

Banerjee, Abhijit, and Drew Fudenberg, "Word of Mouth Learning," working paper, MIT, 1995.

Berkowitz, Leonard, "Studies of the Contagion of Violence." In Herbert Hirsch and David C. Perry, eds., *Violence as Politics: A Series of Original Essays.* New York: Harper & Row, 1973.

Bernardo, Anthony, and Ivo Welch, "A Theory of Overconfidence and Entrepreneurship," working paper, UCLA Anderson School, 1997.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy,* 1992, *100,* 992–1026.

Cao, Henry, and David Hirshleifer, "Limited Observability, Reporting Biases, and Informational Cascades," working paper, University of Michigan, 1997a.

Cao, Henry, and David Hirshleifer, "Word of Mouth Learning and Informational Cascades," working paper, University of Michigan, 1997b.

Caplin, Andrew, and John Leahy, "Business as Usual, Market Crashes, and Wisdom after the Fact," *American Economic Review,* 1994, *84,* 548–65.

Chamley, Christophe, and Douglas Gale, "Information Revelation and Strategic Delay in Irreversible Decisions," *Econometrica,* 1994, *62,* 1065–85.

Chaudhuri, Shubham, Jith Jayartne, and Angela Chang, "Informational Externalities and the Branch Location Decisions of Banks: An Empirical Analysis," working paper, Columbia University, 1997.

Choi, Jay Pil, "Herd Behavior, the 'Penguin Effect', and Suppression of Information Diffusion: An Analysis of Informational Externalities and Payoff Interdependency," *Rand Journal of Economics,* 1997, *28,* 407–25.

"Did Dirty Tricks Create a Best-Seller?" *Business Week,* August 7, 1995, 22.

DiIulio, John J., "Help Wanted, Economics, Crime and Public Policy," *Journal of Economic Perspectives,* Winter 1996, *10*:1, 3–25.

Gibson, Robert M., and Jacob Hoglund, "Copying and Sexual Selection," *TREE,* July 1992, *7*:7, 229–32.

Gilbert, R. J., and Marvin Lieberman, "Invest-

ment and Coordination in Oligopolistic Industries," *Rand Journal of Economics,* 1987, *18,* 17–33.

Glaeser, Edward, Bruce Sacerdote, and Jose Scheinkman, "Crime and Social Interactions," *Quarterly Journal of Economics,* 1996, *111,* 507–48.

Gul, Faruk, and Russell Lundholm, "Endogenous Timing and the Clustering of Agents' Decisions," *Journal of Political Economy,* 1995, *103,* 1039–66.

Hendricks, Kenneth, and Dan Kovenock, "Asymmetric Information, Information Externalities, and Efficiency: The Case of Oil Exploration," *Rand Journal of Economics,* 1989, *20,* 164–82.

Hirshleifer, David, and Noah, Robert, "Misfits and Social Progress," working paper, University of Michigan, 1997.

Hoffer, Eric, *The Passionate State of Mind,* aphorism 33. New York: Harper, 1955.

Huff, C. Ronald, "Denial, Overreaction, and Misidentification: A Postscript on Public Policy." In C. Ronald Huff, ed., *Gangs in America,* Newbury Park, CA: Sage Publications, 1990.

Kahan, Marcel, "Social Influence, Social Meaning, and Deterrence," *Virginia Law Review,* 1997, *83,* 276–304.

Kennedy, Robert E., "Strategy Fads and Strategic Positioning: An Empirical Test for Herd Behavior in Prime-Time Television Programming," Harvard Business School, Division of Research, Working Paper, 1997.

Khanna, Noreen, "Optimal Contracting with Moral Hazard and Cascading," *Review of Financial Studies,* 1998, *11,* 557–94.

Klein, Daniel B., and John Majewski, "Plank Road Fever as Informational Cascade: The Importance of Revelation," working paper, University of California, Irvine, 1996.

Kuran, Timur, "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution," *Public Choice,* 1989, *61,* 41–74.

Landes, William, "An Economic Study of U.S. Aircraft Hijacking, 1961–76," *Journal of Law and Economics,* 1978, *21,* 1–32.

Lee, In Ho, "On the Convergence of Informational Cascades," *Journal of Economic Theory,* 1993, *61,* 396–411.

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

Machiavelli, Niccolo, *The Prince* (1514). Quentin Kinner and Russel Price, ed., Cambridge: Cambridge University Press, 1988.

Miller, Walter B., "Why the United States Has

Failed to Solve Its Youth Gang Problem." In C. Ronald Huff, ed., *Gangs in America.* Newbury Park, CA: Sage Publication, 1990.

"The New Hucksterism: Stealth Ads Creep into a Culture Saturated with Logos and Pitches," *Business Week,* July 1, 1996, 76f.

Office of Juvenile Justice and Delinquency Prevention, Department of Justice, "Gang Suppression and Intervention: Problem and Response," October 1994.

Perktold, Josef, "Recurring Informational Cascades," working paper, University of Chicago, 1996.

Phelps, Charles E., and Cathleen Mooney, "Variations in Medical Practice Use: Causes and Consequences." In Arnold Richard, Robert Rich, and William White, eds., *Competitive Approaches to Health Care Reform,* Washington D.C., Urban Institute, 1993, ch. 7, 140–78.

Robin, Eugene D., *Matters of Life and Death: Risks vs. Benefits of Medical Care.* New York: Freeman and Co., 1984.

Rogers, Everett M., *Diffusion of Innovation,* 3rd Edition. New York: Free Press, Macmillan Publishers, 1983.

Shefrin, Steven M., and Robert K. Triest, "Can Brute Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance." In Joel Slemrod, ed., *Why People Pay Taxes,* 1992, 193–218.

Skogan, Wesley G., *Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods,* New York: Free Press, 1990.

Smith, Lones, and Peter Sorenson, "Pathological Outcomes of Observational Learning," working paper, MIT, 1995.

"A Star is Reborn," *Business Week,* July 8, 1996, 102–6.

Taylor, Richard, *Medicine Out of Control: The Anatomy of a Malignant Technology.* Melbourne: Sun Books, 1979.

Welch, Ivo, "Sequential Sales, Learning and Cascades," *The Journal of Finance,* 1992, *47,* 695–732.

Vives, Xavier, "How Fast Do Rational Agents Learn?" *Review of Economic Studies,* 1993, *60,* 329–47.

Zhang, Jianbo, "Strategic Delay and the Onset of Investment Cascades," *Rand Journal of Economics,* 1997, *28,* 188–205.