

JOSEPH HENRICH

Department of Anthropology

Emory University

and

Wissenschaftskolleg zu Berlin

Wallotstrasse 19

14193 Berlin, Germany

## Cultural Transmission and the Diffusion of Innovations: Adoption Dynamics Indicate That Biased Cultural Transmission Is the Predominate Force in Behavioral Change

---

In challenging the pervasive model of individual actors as cost-benefit analysts who adapt their behavior by learning from the environment, this article analyzes the temporal dynamics of both environmental (individual) learning and biased cultural transmission processes by comparing these dynamics with the robust "S-shaped" curves that emerge from the diffusion of innovations literature. The analysis shows three things: (1) that environmental learning alone never produces the S-shaped adoption dynamics typically observed in the spread of novel practices, ideas, and technologies; (2) that biased cultural transmission always produces the S-shaped temporal dynamics; and (3) that a combination of environmental learning and biased cultural transmission can generate S-dynamics but only when biased cultural transmission is the predominate force in the spread of new behaviors. These findings suggest that biased cultural transmission processes are much more important to understanding the diffusion of innovations and sociocultural evolution than is often assumed by most theorists. [*diffusion of innovations, cultural transmission, learning, cultural evolution*]

---

Efforts to understand human behavioral change have produced a multiplicity of different approaches. Most of these, whether they come from anthropologists, economists, sociologists, or political scientists, share one common core element: the notion that individuals choose among alternative behaviors by performing cost-benefit analyses using payoff-relevant information (i.e., data about the costs and benefits). Laying aside the hyper-rational, omniscient beings of classical economics formations,<sup>1</sup> more plausible approaches model individuals as goal seekers with limited computational abilities and incomplete information, who rely on trial-and-error learning, experimentation, and long experience in similar environments to achieve locally effective solutions (Camerer and Ho 2000; Chibnik 1981; Earle 1997; Erev and Roth 2000; Gladwin and Butler 1984; Harris 1979; Netting 1993; Young 1998). Using data from the vast diffusion of innovations literature, I argue that human behavioral change does not result primarily from individual-level trial-and-error learning or cost-benefit analysis. Instead, I show that the dynamics of diffusion demand a primary reliance on some form of biased cultural transmission.

The persuasive model of human behavior that pervades the social sciences proposes that individuals acquire and evaluate *payoff-relevant* information about alternative be-

havioral options by action and interaction in their local social, economic, and ecological environments (Camerer and Ho 2000; Chibnik 1981; Earle 1997; Erev and Roth 2000; Gladwin and Butler 1984; Harris 1979; Young 1998). The adjective *payoff-relevant* emphasizes that the information analyzed by individuals is directly applicable to evaluating behavioral alternatives, according to some set of prescribed goals. These goals may involve concepts such as self-interest, reproductive fitness, social prestige, income, and group benefits. Here, I argue against this standard model by showing three things. First, environmental learning models alone, without substantial contributions from biased cultural transmission, do not generally produce the empirical "S-shaped" cumulative adoption curves that dominate the diffusion of innovations literature. Second, biased transmission models alone, and especially those with a *conformist transmission* component, consistently produce the particular S-dynamics found throughout the literature. And, third, a combined model, with both environmental learning and biased cultural transmission, allows us to predict the conditions that produce the different kinds of empirically observed diffusional dynamics. However, this model only generates S-dynamics when biased transmission predominates.

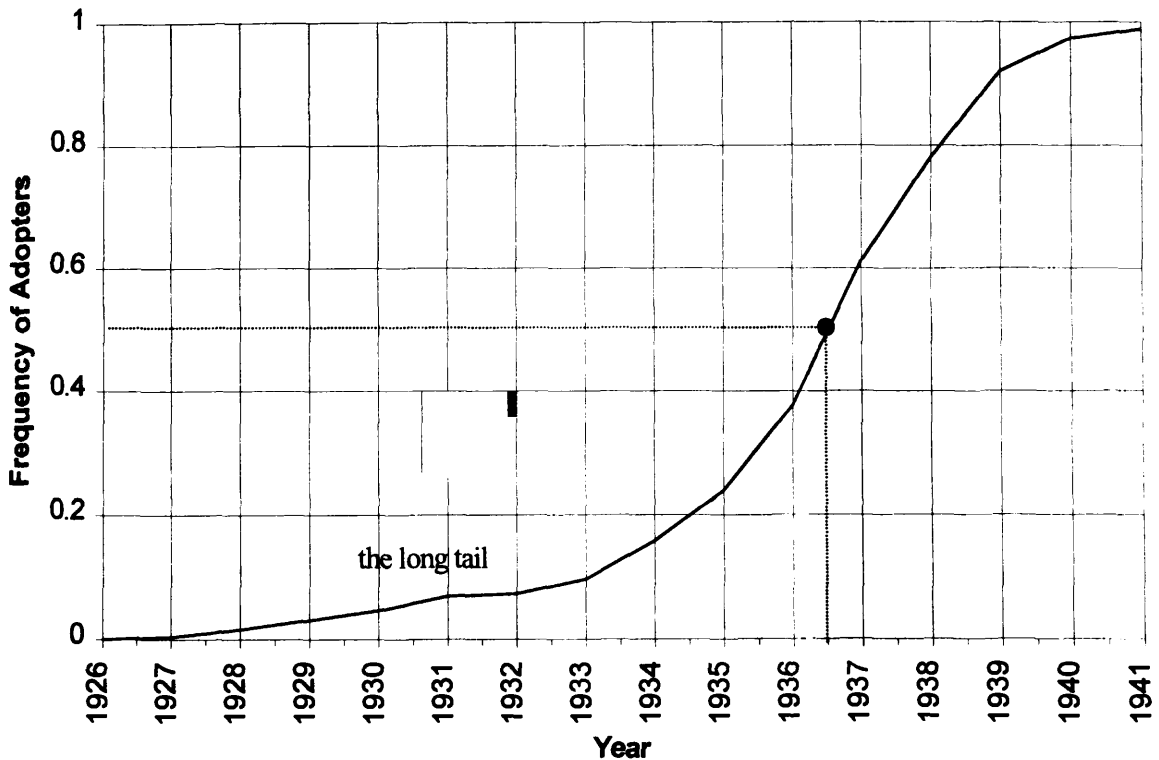
### S-Shaped Adoption Curves

One of the most robust findings from over 3,000 studies in the diffusion of innovation literature is the S-shaped cumulative adoption curve (Rogers 1995:23). This vast literature contains data for the spread of an enormous variety of practices, technologies, and ideas in communities and countries throughout the world. These cases include the adoption of “innovations” such as hybrid corn among Iowa farmers, bottle-feeding practices among impoverished Third Worlders, new governance practices among Fortune 500 companies, chemical fertilizers among small-scale farmers, novel approaches to mathematics training (the “new math”) among secondary schools, and the practice of not smoking among Americans. Typically, the cumulative adoption curve for the spread of these practices has an S-shape. For example, Figure 1 shows the S-curve that emerges from Ryan and Gross’s classic study (1943) of the spread of hybrid corn in two Iowa farming communities. This general shape captures the temporal dynamics encountered in a wide range of diffusion studies.

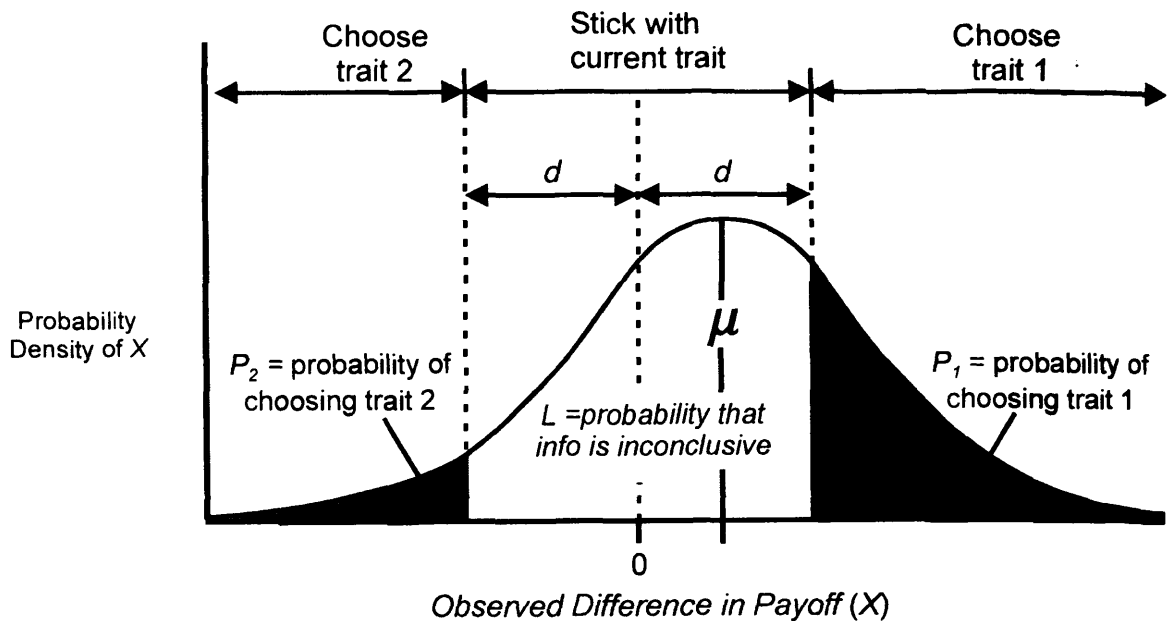
However, not all adoption curves are S-shaped. Of the small fraction of curves that are not S-shaped, most display

a single alternative shape, which I will call an *R*-curve. *R*-curves lack the slow growth during the initial portion of the spread, which characterizes *S*-curves (the bottom left part of the curve in Figure 1). Instead, *R*-curves begin at their maximum rate of growth (at  $t = 0$ ) and then slowly taper off toward equilibrium. *R*-curves describe, for example, the cumulative adoption dynamics for the spread of the practice of prescribing Tetracycline among “isolated” doctors (Coleman et al. 1966). They also characterize the spread of milk bottle-opening behaviors among pigeons (Lefebvre and Giraldeau 1994) and the spread of potato washing among Japanese macaques (Kawai 1965; Tomasello 2000). Interestingly, *R*-curves also arise from a variety of nonsocial learning processes in which individuals (humans) acquire increasing proficiency in some skill or ability through practice (see Jovanovic and Nyarko 1995).

To explore the relative importance of environmental (individual) learning versus biased cultural transmission, I have analyzed the dynamics of three models: a generalized individual-level environmental learning model, a biased cultural transmission model, and a combined model. The combined model integrates the first two models and allows



**Figure 1.** Diffusion of hybrid corn seed in two Iowa farming communities. This diffusion curve provides a prototypical example of a “long-tailed” *S*-curve. The dotted line marks the point on the curve with the highest rate of change ( $dq/dt$ ). Adapted from Ryan and Gross 1943.



**Figure 2.** The environmental learning model. In this model, each individual receives an independent “signal” (a piece of information) from the environment (perhaps via experiments) about the difference in payoffs between two alternative behaviors. This signal is a random draw from the above probability distribution, which has mean  $\mu$  and standard deviation  $\sigma = 1$ . The larger  $\mu$  is, which is the average (and modal) difference in payoffs indicated by the signal, the more clearly this signal distinguishes Trait 1 from Trait 2. Symbol  $d$  sets an individual’s threshold of evidence for adopting a trait—that is, how clear the information must be before the individual will switch. This means that  $P_1$  gives the individual’s probability of adopting Trait 1,  $P_2$  gives their probability of adopting Trait 2, and  $L$  gives their probability of not changing ( $L$  gives their probability of relying on unbiased transmission). This figure is adapted from Henrich and Boyd 1998.

us to compute the relative contribution of each to the *S*- and *R*-adoption dynamics.

### The Environmental Learning Model

Figure 2 graphically depicts a simple, though quite general, model of environmental learning for two behavioral traits.<sup>2</sup> In the typical diffusion of an innovation, tracking only two traits is sufficient to capture the essential process: Trait 1 represents the presence of the novel trait (the “innovation”), and Trait 2 indicates the absence of the trait. If we are, for example, studying the spread of a new nitrogen fertilizer, an individual possesses Trait 1 if he or she uses the fertilizer and possesses Trait 2 if he or she does not use the fertilizer. The normal curve in Figure 2, with mean  $\mu$  and variance  $\sigma^2$ , shows the distribution of relative payoff information provided by the environment. Individuals may acquire this information through observation, experience, interaction, or experimentation in the environment. For simplicity, we will set  $\sigma^2$  equal to 1 and refer only to effects of changing  $\mu$ . During each time cycle (a fixed time period), each individual in the population receives one independent draw from this normal distribution. This

single draw provides a measure of the difference in payoffs  $X$  between the two alternative behaviors. However, people do not switch to a novel behavior based only on one piece of information, unless the suggested payoff difference is sufficiently large. How large this value of  $X$  needs to be depends on the quality of environmental information available, which is captured by  $\mu$ , and on the individual’s “threshold of evidence”—which is parameterized by  $d$ . If, for example, the  $X$  drawn during a given cycle falls between  $-d$  and  $+d$ , the individual stays with his or her previous behavior (from the previous time cycle). However, if the  $X$  drawn exceeds  $d$ , then the individual switches to Trait 1. If people already possess Trait 1, they stick with it. If  $X$  falls below  $-d$ , then the individual switches, mistakenly, to behavioral Trait 2 or retains it if he or she already has it. This is a “mistake” because the situation depicted in Figure 2, by the fact that  $\mu > 0$ , indicates that behavioral Trait 1 is superior in the current environment. *Superior* means that Trait 1 brings higher payoffs, on average, relative to whatever individuals want, strive for, or hope to maximize.

To illustrate this phenomenon, suppose a farmer, who currently plants wheat variety *A*, decides to plant a small patch of land with a novel wheat seed, variety *B*, as an experiment.<sup>3</sup> This experimental patch provides our farmer with a single measure of average yield (in kilograms of wheat harvested per hectare, for example), which we can compare against the average yield for variety *A*. The difference between the yield per hectare for the experimental patch and the average yield per hectare for variety *A* provides a value of *X*—an observed difference in payoffs between the two varieties. If the yield from variety *B* is about the same as or less than that from variety *A* (implying  $X < d$ ), then our farmer does not change from variety *A*. However, if the yield from variety *B* is sufficiently greater than the yield from *A* ( $X > d$ ), then our farmer switches and sows only variety *B* in the following year. Note, for this illustration we assume that all other aspects of *A* and *B* (besides yield) that might concern a farmer are identical.

Now, let us derive the population dynamics for the spread of Trait 1 into a group in which everyone currently possesses behavioral Trait 1. The symbol *q* gives the frequency of individuals with Trait 1 in a large population, while  $1 - q$  gives the frequency of individuals with Trait 1. Initially,  $q = 0$ , but with each time cycle we update the value of *q*. The new value of *q*, in the next time cycle, is represented as *q'* (which reads “*q* prime”). Applying the environmental learning model described above and depicted in Figure 2 to a large population, we arrive at the following recursion:

$$q' = q + \underbrace{(1 - q)P_1}_{\text{Trait 1 adopters}} - \underbrace{qP_2}_{\text{Trait 2 adopters}} \quad (0)$$

This says that the expected frequency of individuals with Trait 1 in the next time step, *q'*, equals the frequency of individuals with Trait 1 in the previous time step, *q*, plus the proportion of individuals who had Trait 1 but received an *X* value greater than *d* (so they switch to Trait 1) minus that proportion of individuals who had Trait 1 but received an *X* value less than  $-d$  (so they adopt Trait 2). Using the fact that  $L = 1 - P_1 - P_2$  (see Figure 2 and Appendix A), the above equation simplifies to Equation 1:

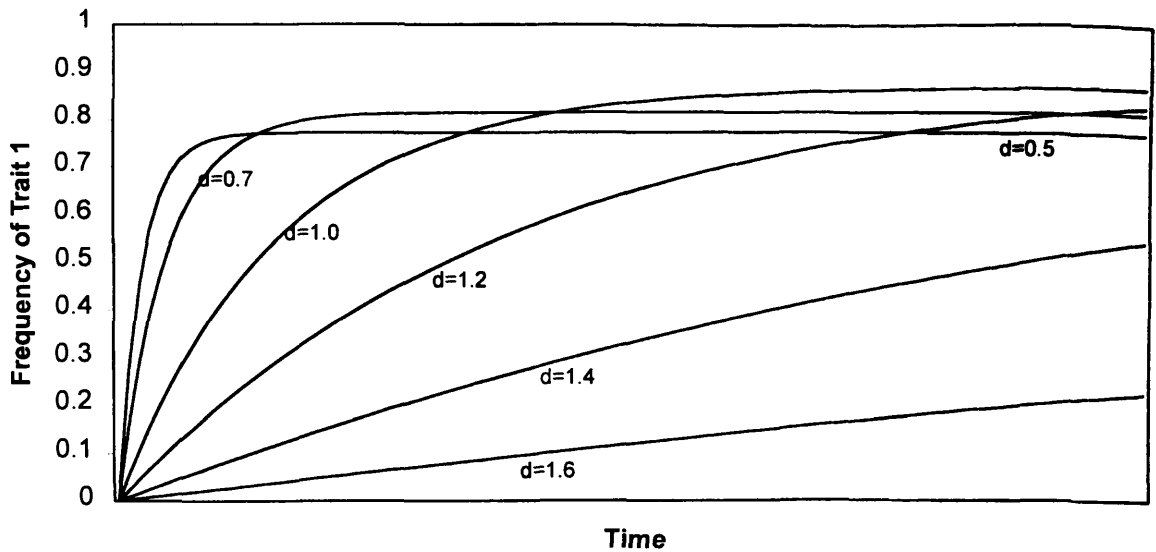
$$q' = P_1 + Lq \quad (1)$$

The new frequency of individuals with Trait 1, *q'*, depends on  $P_1$ , *L*, and *q*.  $P_1$  is the probability of learning the new trait from environmental information obtained during this time cycle. Restated, it is the probability that the payoff difference observed between the two behaviors exceeds the threshold of evidence (*d*)—it is also the gray area under the curve on the right side of Figure 2. *L* is the probability that the environmental information is inconclusive and represents the area between  $-d$  and  $+d$  under the curve in Figure 2. Individuals who receive inconclusive information will stick with their current behavior. Both  $P_1$  and *L*

are derived from *d* and  $\mu$  via the distribution shown in Figure 2—Appendix A outlines this derivation. By iterating Equation 1 recursively through successive time cycles, we can plot its temporal dynamics and the cumulative adoption curves that it generates. Figure 3 shows these dynamics for a series of curves generated by Equation 1 for different values of *d* (with a fixed value of  $\mu$ ).

For those readers who are, like me, interested in longer term cultural evolutionary processes—in which the frequency of different ideas, beliefs, values, and practices may change over many generations—we can interpret Equation 1 in a slightly different way. Suppose that during each time cycle, or perhaps each generation, naive individuals (those who do not currently possess a particular behavior) first acquire a behavior (Trait 1 or 2) by unbiased transmission—that is, by copying their parents or other individuals at random from the population. This form of social learning is called unbiased because, at the population level, it simply replicates the distribution of behaviors from the previous generation,  $q' = q$ . Then, after acquiring a behavior, individuals obtain environmental information about the relative payoffs of alternative behaviors (a value of *X*). If the difference in payoffs is clear (that is, if *X* is greater than  $+d$  or less than  $-d$ ), then individuals adopt the behavior indicated by their environmental information. However, if *X* falls between  $+d$  and  $-d$ , individuals stick with the behavior they acquire via unbiased cultural transmission. Because the driver of change in this model lies entirely in the evaluation of environmental information, this setup also yields Equation 1. Boyd and Richerson (1985, 1988) call this combination of unbiased transmission and (“individual”) environmental learning *guided variation*. Because guided variation explicitly includes unbiased transmission, I have called this the “environmental learning model” to include both the individual-level learning process I have described (which may occur many times per generation) and its transgeneration counterpart, guided variation (individual learning plus unbiased transmission). We need these two conceptualizations to deal with both diffusions that occur within a generation and longer term diffusions that occur over several generations.

The environmental learning model (usually in its guided variation form) captures the fundamental processes that many economically oriented anthropologists believe underlie much of sociocultural evolution. From this perspective, cultural evolution, behavioral change, and adaptation result from individuals making cost-benefit decisions and transmitting these decisions, or their behavioral outputs, via social learning to their progeny. So, although most see “culture” or “tradition” as part of the process, the actual cultural transmission among individuals (or generations) plays no dynamic role. Cultural transmission simply replicates the existing distribution of behaviors, beliefs, and so on (Binford 1983:222) without substantially altering their



**Figure 3.** Environmental learning *R*-curves for different values of *d*. The symbol *d*, shown in Figure 2, sets an individual's "threshold of evidence" for changing behavior (traits). As *d* increases, the time required to reach equilibrium (the point when the frequency stops changing) is longer, but the equilibrium frequency of the favored trait is greater (compare *d* = 1.0 and *d* = 0.5). Note that the equilibrium frequencies of the curves *d* = 1.2, 1.4, and 1.6 are not shown—but these are greater than all of the other curves. This plot illustrates the tradeoff in environmental learning between the speed of adoption and the accuracy of adoption.

distribution or form—that is, unbiased transmission. The driver of changes lies in the decision-making process, not in the transmission process. Gladwin and Butler (1984: 210), following Chibnik (1981), articulate this approach in three steps: (1) individuals evaluate alternatives using low-cost experiments to gather information, (2) these decisions become codified in cultural rules, and (3) these rules are transmitted (unbiased) to the next generation. For Gladwin, Butler, and Chibnik, the driver of behavioral change lies in the cost-benefit evaluation of alternatives based on low-cost experimentation, not in the transmission of this information among individuals.

Similarly, while Harris maintains that "as a species we have been selected for our ability to acquire elaborate repertoires of socially learned responses" (1979:62), he believes that sociocultural evolution is driven by individuals opportunistically selecting among cultural/behavioral variants according to their cost-benefit ratios. The second assertion about cost-benefit ratios can only be true if the apparent social learning abilities of humans do not substantially bias the intergeneration transmission of cultural/behavioral variants. Tomasello (2000:38) lays out similar sequences of individual learning and unbiased transmission in his recent book on human cognition and culture (also see Boyd and Richerson 1985: ch. 4). Such a process can also be gleaned from Binford (1983). Consequently, the position of Harris, Gladwin, Tomasello, Bin-

ford, Chibnik, and Butler, that sociocultural evolution results from unbiased social learning plus opportunistic cost-benefit analysis (environmental learning), is exactly what guided variation attempts to formalize. With this formalization, we can now better analyze its evolutionary dynamics. If the empirical data show that cultural transmission biases (specified below) do substantially affect the frequency of alternative cultural/behavioral variants from one generation to the next,<sup>4</sup> then the environmental learning approach fails to capture an important component of sociocultural evolution.

Using data from the diffusion of innovations literature, we can address the applicability of the environmental learning models. If the evaluation of costs and benefits, based on environmental information, is the dominant force in the spread of novel practices, then empirically observed cumulative adoption curves should reveal the basic *R*-shape generated by Equation 1 and shown in Figure 3. Interestingly, however, most adoption curves constructed from empirical data have the *S*-shape shown in Figure 1, not the *R*-shape seen in Figure 3. An examination of empirical *S*-curves tells us that the change in *q* over each time cycle must first increase to a maximum point somewhere in the middle of the *S* and then begin decreasing toward 0. Computing  $\Delta q$ , the change in *q* over each time cycle, we get the following:

$$\Delta q = q' - q = P_1 - q(1 - L) \quad (2)$$

Note that both  $P_1$  and  $1 - L$  are positive constants, so  $\Delta q$  must decrease as  $q$  increases. Consequently, Equation 1 will never produce an S-shape.

Although R-curves do occasionally appear in the diffusion literature, S-curves are, by far, the dominant shape of the temporal dynamics. Therefore, either this general environmental learning model somehow fails to capture the logic of humans as cost-benefit decision makers or behavioral change is *not* primarily driven by individual-level environmental learning. With this, it is important to realize that, even if people from different social groups have different “logics” for assigning and evaluating costs and benefits (varying  $\mu$  and  $d$  changes the perceived payoffs), the temporal dynamics will remain R-shaped. So, although some groups may have different R-shapes (or lack diffusions at all), the above conclusion does not change. Later, after I have presented the basic biased cultural transmission model and the combined model, I will modify this environmental learning model and add the assumption that individuals vary in their degree of “innovativeness.” As we will see, this modification does not change the basic results just derived.

### The Biased Cultural Transmission Model

Instead of assuming that individuals acquire novel traits by figuring things out, using payoff-relevant information acquired directly from the environment, a substantial amount of empirical work from throughout the social sciences suggests that humans rely on social learning or cultural transmission to acquire the majority of their behaviors (Bandura 1977; Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981; for a summary, see Henrich 2002). However, people do not simply imitate random things from random people. Rather, the social learning process seems to contain a number of biases. Here, I will describe three categories of transmission biases: *direct bias*, *prestige bias*, and *conformist bias* (biased cultural transmission is also termed “cultural selection”; see Durham 1991:192, 1992:334).

Direct biases result from cues that arise from the interaction of specific qualities of an idea, belief, practice, or value with our social learning psychology. The presence of these cues makes people more (or less) likely to acquire a particular trait. Evolutionary anthropologists have theorized that natural selection favors a variety of such biases in our psychology because they promote the acquisition of fitness-enhancing cultural traits via social learning (Boyd and Richerson 1985; Durham 1991; Richerson and Boyd 1992). However, in thinking about direct bias it is important to keep in mind two things. First, jury-rigged evolutionary products, like human minds, are likely to contain lots of accidental by-products and latent structures that create biases among fitness-neutral behaviors, ideas, beliefs, and values. Second, direct biases that led to the adoption of

fitness-enhancing behavior in ancient environments (like a bias toward acquiring the practice of eating certain salty, fatty foods) may now promote the adoption of quite maladaptive practices. For example, the practice of purchasing and using cooking oil spreads rapidly even through remote villages—far from the reach of advertising—because there is something about the behavior or idea that appeals to people.<sup>5</sup>

Under prestige-biased transmission, people copy ideas or practices from *individuals* with specific qualities or attributes, regardless of the characteristics of the behaviors or ideas that are copied. Gil-White and I have demonstrated that people will copy a wide range of traits from prestigious or successful people, even when the behaviors, ideas, or opinions have nothing to do with the person’s prestige or success (Henrich and Gil-White 2001).<sup>6</sup> Americans, for example, will use a certain type of cologne or even shave their heads if Michael Jordan does (or they believe he does), despite that fact that Jordan’s scent and hairstyle are probably not connected to his basketball prowess, prestige, and overall success.

Finally, under conformist transmission, humans preferentially imitate ideas and behaviors that are expressed by a majority of the group over traits expressed by the minority, even when their personal opinions or behavior will *not* be known by the other group members (Baron et al. 1996; Bettinger et al. 1996:151; Insko et al. 1985; for theoretical treatment, see Henrich and Boyd 1998). All three of these transmission biases rely on cues that may, in particular cases and perhaps on average in some ancestral human environment, arise indirectly from the difference in fitness payoffs between alternative behaviors ( $\mu$  in Figure 2). But in more recent human environments there is no reason to assume any consistent connection; even if there were such a consistent connection, the temporal dynamics of diffusion driven by environmental learning will still look quite different than those produced by biased transmission.

Equation 3 formalizes biased cultural transmission and was derived using basic replicator dynamics (Boyd and Richerson 1985; Weibull 1995). As in Equation 1,  $q$  represents the frequency of individuals with the novel behavioral trait (Trait 1), and  $q'$  is the frequency of individuals with Trait 1 in the next time cycle (Appendix A supplies a simple derivation):

$$q' = q + (1 - q)q(r_1 - r_2) = q + q(1 - q)B \quad (3)$$

The term  $r_1 - r_2$ , or simply  $B$ , ranges from  $-1$  to  $1$  and represents the overall difference in the replicatory propensities of Traits 1 and 2. These replicatory propensities may depend on particular characteristics of Trait 1 and/or 2 (direct bias) that may or may not be indirectly linked to payoff differences. Or they may depend on the prestige (or success) of the individual(s) currently possessing the trait (prestige bias), which also may or may not be an indirect measure of payoffs. Or they may depend on the current frequency of

individuals possessing the trait (conformist transmission), which may or may not be indirectly linked to payoffs. For now, we will leave conformist transmission out and assume that  $B$  aggregates only the effects of prestige and direct bias. We will also assume that these biases are not a function of either  $q$  or time.<sup>7</sup> Later, I will incorporate conformist transmission (which is frequency dependent) and examine its influence on diffusion. The reader should be aware that Equation 3 is not some special case of cultural transmission but, rather, a general form for any replicator process. It has been independently rederived for a variety of purposes in a number of different fields—including economics (Gintis 2000; Weibull 1995), genetics (Hartl and Clark 1989), epidemiology (Waltman 1974), and cultural transmission (Bowles 1998; Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981)—and is often applicable even when traits are continuous (Henrich and Boyd 2001).

Intuitively, we can understand Equation 3 by considering a large population in which each individual encounters one potential “cultural model” (i.e., another person whom one might learn from) during each time cycle, which could be one year, one day, one month, and so on. Under prestige-biased transmission, this individual compares his or her success (e.g., in wealth, number of admirers, or valued skills) to the success of the potential cultural model. If the model has greater success and a different trait than the individual, the individual switches and adopts the trait of the more successful model. If not, the individual keeps his or her original trait. Or, under direct bias, the individual compares the qualities of his or her trait to the qualities of the trait possessed by the cultural model. If the person favors (for whatever psychological reason) the trait possessed by the model over the trait he or she currently possesses, he or she switches and adopts the model’s trait. If we then ask (for either the direct- or the prestige-biased cases), what the expected frequency of one of the traits (i.e., “Trait 1”) after one unit of time is, we arrive at Equation 3. This expression tells how the frequency of the traits will change from one time step to another because of the social learning processes described above. As I have said, this intuitive layout can be modified in a variety of sensible (and silly) ways that all yield Equation 3 or something quite similar (Appendix A formalizes this).

Figure 4 presents four cumulative adoption curves generated using Equation 3 for different values of  $B$ . Note the similarity between the empirical curve in Figure 1 and the curves in Figure 4. In fact, the different  $S$ -shapes captured in Figure 4 resemble a wide range of the empirical adoption curves found in the diffusion of innovations literature. This similarity suggests that cultural transmission models may capture an important component of human behavioral change.

## The Combined Model: Environmental Learning + Biased Transmission

So far, I have contrasted two quite different models of human learning. However, it seems both intuitively and empirically true that humans do both biased cultural transmission and environmental learning. That is, we do some selective imitating and some figuring things out on our own. Boyd and I have developed this idea theoretically, using computer simulations that model the biological evolution of the parameter  $d$ —which, as I discussed earlier, determines an individual’s degree of reliance on environmental learning versus cultural transmission (Henrich and Boyd 1998). Under a wide range of conditions, in both spatially and temporally varying environments, this theoretical work suggests that our reliance on environmental learning is a small, but important, component of human adaptive behavioral plasticity. Consequently, the question becomes, How much biased cultural transmission must be added to environmental learning (or the guided variation model) to generate the empirically observed  $S$ -curves? Or what is the predominate force in human behavioral change?

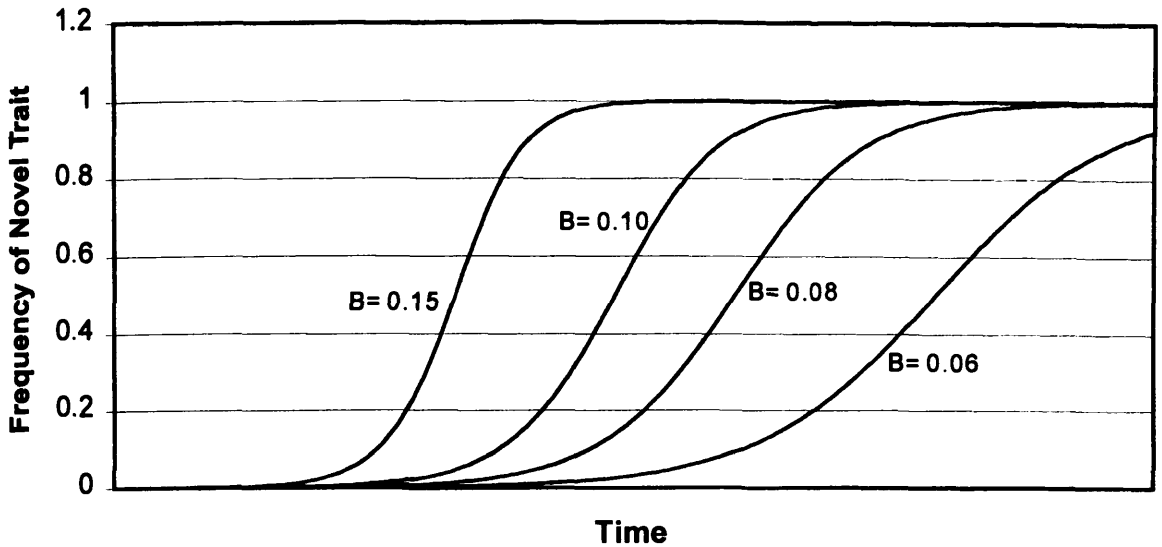
To address this, we combine Equations 1 and 3. However, because simply substituting Equation 1 into Equation 3 gives a slightly different answer than substituting Equation 3 into Equation 1,<sup>8</sup> an additional step is required. We assume that, during each time cycle, not everyone attempts environmental learning or biased transmission. Instead, only a fraction of the population updates behavior based on one of these two sources of information. For environmental learning, the symbol  $\xi$  represents the fraction of individuals in the population who consider updating their behavior via environmental learning per unit of time. This can be thought of as the update rate for environmental learning or as the probability of using environmental learning in each time cycle. Similarly, the symbol  $\gamma$  represents the fraction of individuals in the population who update with biased cultural transmission per unit of time. In both cases,  $\Delta t$  represents one unit of time or one time cycle. Therefore,  $\xi\Delta t$  provides the fraction of individuals who consider updating with environmental learning in each time cycle, while  $\gamma\Delta t$  gives the fraction of individuals who deploy biased transmission in a given time cycle. Applying this additional step to Equation 1 yields the following:

$$q'_L = q(1 - \xi\Delta t) + (P_1 + Lq)\xi\Delta t = q + \xi\Delta t(P_1 + [L - 1]q) \quad (4)$$

Applying the same process to Equation 3 yields Equation 6:

$$q'_T = q(1 - \gamma\Delta t) + \gamma\Delta t(q + q[1 - q]B) = q + \gamma\Delta t(q + q[1 - q]B) \quad (5)$$

Because we want to arrive at the derivative of  $q$  with respect to time, we substitute Equation 4 into Equation 5 (or Equation 5 into Equation 4), solve for  $\Delta q/\Delta t$ , and then take the limit of  $\Delta q/\Delta t$  as  $\Delta t$  approaches 0. This gives us Equation 6, which describes the rate of change in the frequency of Trait 1 as a function of  $q$ :



**Figure 4.** Biased cultural transmission dynamics using four different values of  $B$ . As  $B$  decreases, the amount of time to equilibrium increases, but the final equilibrium value remains 1 (everyone adopts).

$$\frac{dq}{dt} = \xi(P_1 - [1 - L]q) + \gamma Bq(1 - q) \quad (6)$$

An examination of the typical  $S$ -curve (Figure 1) suggests the general form of  $dq/dt$ . The rate of change of the frequency of the novel trait must first ascend to a peak, somewhere in the middle of the  $S$ -curve, and then decline to 0. Figure 5 plots  $dq/dt$  for both the combined model (CM, Equation 6) and the environmental learning model (ENLR, Equation 4). For ENLR, the maximum value of  $dq/dt$  occurs at  $q = 0$ , so no  $S$ -curve is generated. For the biased cultural transmission model, which is not shown, the curve has the same shape as the CM curves, except the maximum value of  $dq/dt$  (the middle of the  $S$ ) always occurs at  $q = 0.50$ . For the CM (without conformist transmission), the maximum value of  $dq/dt$  occurs below  $q = 0.50$ . When this maximum value occurs between  $q = 0$  and  $q = 0.50$ , some form of  $S$  is produced. When this maximum value occurs below  $q = 0$ , the part of the  $dq/dt$  curve between  $q = 0$  and  $q = 1$  looks just like the environmental learning curves. Consequently, an  $R$ -curve, not an  $S$ -curve, is produced. To visualize this, imagine sliding the CM curve, shown in Figure 5, to the left until the maximum value of  $dq/dt$  drops below 0—the part of the CM curve on the right side of 0 resembles the ENLR curve.

To derive the conditions under which the CM produces an  $S$ -curve, we take the derivative of Equation 6 with respect to  $q$ , set it equal to 0, and solve for  $q$ . This provides an expression for the value of  $q$  when  $dq/dt$  is maximized:

$$q_{\max} = \frac{1}{2} - \frac{(1 - L)\xi}{2B\gamma} \quad (7)$$

In order to produce the  $S$ -shape, mathematically speaking,  $q_{\max}$  must be greater than 0. So, solving Equation 7 for  $q_{\max} > 0$  yields the following:

$$B > \frac{\xi(1 - L)}{\gamma} = \phi(1 - L) \quad (8)$$

Because  $\xi$  and  $\gamma$  are both update rates, we can simplify Equation 8 by defining  $\phi = \xi/\gamma$ , where  $\phi$  represents a ratio of the fraction of the population that updates via environmental learning to the fraction that updates via biased cultural transmission. If the update rates are equal, then  $\phi = 1$ ; if people update their behavior more frequently using environmental information, then  $\phi > 1$ ; if people use cultural transmission more frequently, then  $\phi < 1$ . Figure 6 graphs the  $S$ - and  $R$ -regions of  $B$  and  $L$ . This plot shows that in order to consistently produce  $S$ -curves, either  $B$ , the replicatory bias created by the trait or the individual(s) possessing the trait, must be big, or  $L$ , the degree to which humans rely on cultural transmission over environmental learning, must be big. Figure 6 and Equation 8 provide the minimum mathematical conditions to produce an  $S$ -curve. However, for humans to discern an  $S$ -shape in the curve,  $q_{\max}$  should be set at 0.1 or more. This shrinks the region of  $B$  and  $L$  that generates  $S$ -curves, thus making the argument stronger. Below I discuss why I think it is  $L$  that is consistently big, and not  $B$ .

Figure 7 shows the effect of moving the value of  $\phi$  away from 1. Increasing the update rate of environmental learning relative to cultural updating (i.e., changing  $\phi$  to 1.2) moves the  $B$ -intercept (at  $L = 0$ ) up to 1.2, which shrinks the  $S$ -region—which is the plot area to the right of the



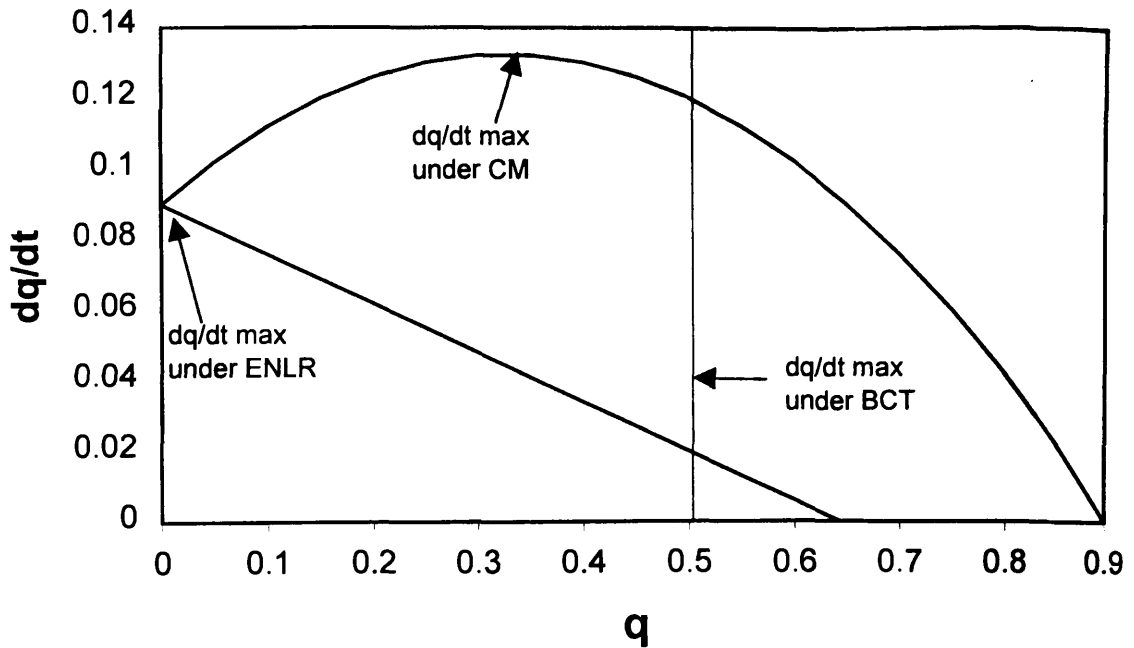


Figure 5. The rate of change of the frequency of the novel trait under the environmental learning model (ENLR) and under the combined model (CM). The vertical line at  $q = 0.50$  shows the  $dq/dt$  maximum value for the biased cultural transmission model (BCT)—otherwise the shape of BCT resembles that of CM.

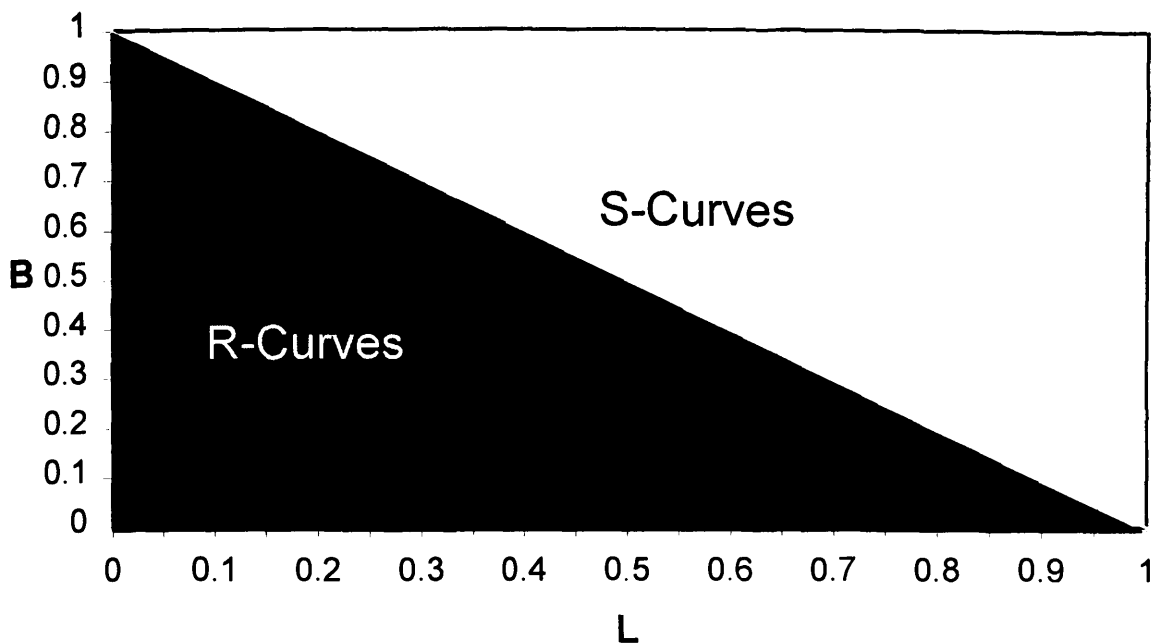
curve. Conversely, increasing the cultural update rate relative to environmental learning (decreasing  $\phi$  to 0.8) moves the  $B$ -intercept down to 0.8 and expands the  $S$ -region. Depending on the details of a particular diffusional situation, one might argue that  $\phi$  is greater or less than 1, but its actual value will be difficult to measure empirically because it depends partly on human psychology and partly on environmental constraints. For much of the coming discussion I assume that  $\phi = 1$ . For empirical purposes, it is best to incorporate  $\gamma$  and  $\xi$  into  $B$  and  $1 - L$ , respectively.

### When Biased Transmission Opposes Environmental Learning

So far, I have considered only the situation in which both biased transmission and environmental learning favor the spread of the novel trait. In this section, I explore the temporal dynamics of the diffusion of a novel trait when individual learning successfully spreads the novel trait against the force of biased transmission—that is, when  $B$  is negative so the transmission bias favors Trait 2. In the next section, I analyze the opposing case, in which biased transmission spreads a novel trait in the face of environmental learning. This occurs when environmental information indicates that a trait is not beneficial but transmission biases spread it anyway. By exploring these two situations, we

can ask which set of dynamics more closely matches the empirically observed temporal dynamics of trait diffusion.

Figure 8 shows the adoption curves for five different sets of parameters ( $B$ ,  $L$ , and  $P$ ). As either  $L$  or  $B$  increases ( $B$  is negative), the equilibrium value of  $q$  rises and the curves ascend more quickly. However, nothing remotely resembling an  $S$ -curve emerges. Equation 8 tells the same story. The right side of Equation 8 is always positive (or 0), and  $B$  is always negative in this case, so the condition in Equation 8 is never satisfied, and  $S$ -curves never emerge. Given that  $S$ -curves are empirically rampant in diffusional contexts, the situations in which individual-level environmental learning overpowers biased cultural transmission to spread a beneficial trait must be relatively rare. Two possible explanations for this present themselves. One suggests that our database is somehow biased against these kinds of diffusions, so they only seem rare—somehow these kinds of diffusions are systematically omitted from study. The second explanation is that  $L$  (or more precisely  $d$ ) is large—meaning that biased transmission is the predominate component of human cognition. In the future, researchers should look for diffusional cases in which trial-and-error learning clearly favors one trait while transmission bias favors another (e.g., when only low-status people initially adopt a trait like boiling drinking water).



**Figure 6.** The regions of  $B$  and  $L$  that produce  $S$ - and  $R$ -dynamics. This plot assumes that the update rates for environmental learning and cultural transmission are equal,  $\phi = 1$ . This plot suggests that  $L$  (or  $d$  via Figure 2) is probably large, implying that unless payoff information is very clear, people rely on biased cultural transmission.

### When Environmental Learning Opposes Biased Transmission

What do the curves look like when biased transmission swims upstream against environmental learning? Examining Equation 8 alone suggests that  $S$ -curves may or may not be generated, depending on the values of  $L$  and  $B$ . This condition holds even when environmental information does not favor the novel trait. Because the bias ( $B$ ) must overcome environmental learning, we can use Equation 8 to set  $B$  to the maximum value that still produces  $R$ -curves,  $B = \phi(1 - L)$ . If we substitute this into Equation 6, set it equal to 0, and solve for  $q$ , we get Equation 9:

$$q_{\text{eq}} = \sqrt{\frac{P_1}{(1-L)}} = \sqrt{\frac{P_1}{P_1 + P_2}} \quad (9)$$

This is the equilibrium frequency of the novel trait when an  $R$ -curve is produced by biased transmission flowing upstream against environmental learning. Remember, in this case  $P_1$  is less than  $P_2$  because environmental information opposes the spread of the trait. Consequently, the frequency produced by Equation 9 is small. Typically, this equilibrium value is so small that it would never “count” as diffusion. Figure 9 shows eight curves for differing values of  $B$ ,  $L$ , and  $P_1$  that illustrate the basic point. All diffusions that lead to high equilibrium frequencies of Trait 1, and are

driven by biased transmission, generate  $S$ -curves, not  $R$ -curves. Curve 8, the only  $R$ -curve in Figure 9, illustrates the case when  $B = \phi(1 - L)$ . The equilibrium frequency of Curve 8 is 0.32.

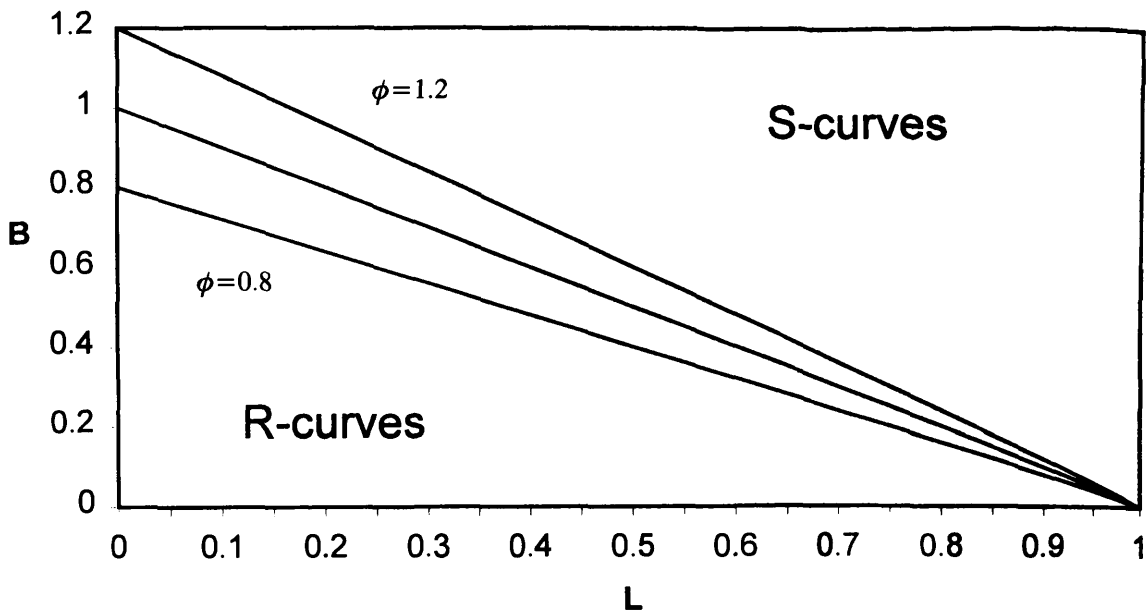
### Conformist Transmission and Long Tails

Figure 1 displays an interesting feature that suggests another form of biased cultural transmission—conformist transmission—may also be at work. Note the slow growth of  $q$  during the initial stages of the diffusional process—I call this slow growth a “long tail” (Figure 1). It took nine years for the frequency of hybrid planters to reach 0.20 but only six more years for it to reach fixation at 0.99. To account for this recurrent phenomenon of long tails, we can incorporate a simple conformist component into the existing model and then examine its effects on the temporal dynamics of adoption.

So far, we have dealt with  $B$ , the replicatory or transmissive bias on the novel trait, as a constant in any particular situation, not as a function of time or frequency. Now  $B$  has two components, a constant part and a frequency-dependent part, which are shown in Equation 10:

$$B = b(1 - \alpha) + \alpha(2q - 1) \quad (10)$$

The second term in Equation 10,  $\alpha(2q - 1)$ , is the component of the overall bias contributed by conformist transmission.



**Figure 7.** *S*- and *R*-curve *B*-*L* regions for different values of  $\phi$ . As  $\phi$  increases, the size of *B*-*L* regions that produces *S*-curves decreases. This means that if people update via environmental learning more frequently than via cultural transmission, then the *B*-*L* combinations that produce *S*-curves are fewer.

The symbol  $\alpha$ , which varies between 0 and 1, gives the relative strength of conformist transmission in human cognition—it scales the cognitive weight given to the frequency of a behavior *relative to* other biases. Generally, it is best to consider  $\alpha$  small, for when  $\alpha$  is large, few if any traits can spread—for example, when  $\alpha > 0.5$ , nothing rare ever spreads. The term  $2q - 1$  varies between  $-1$  and  $1$ . When the frequency of the novel trait is low (less than 50 percent), this conformist component is negative, which reduces the value of the total bias ( $B$ ) and may actually make it negative (depending on the relative sizes of the other components). When  $q > 0.50$ , this conformist term increases the overall size of the bias. The other term,  $b(1 - \alpha)$ , is the contribution to the overall bias made by non-frequency-dependent direct biases and prestige biases. The symbol  $b$  is the constant bias (which can vary from  $-1$  to  $1$ ), while  $1 - \alpha$  gives the weight accorded to the nonconformist component of the transmission bias. Substituting Equation 10 into Equation 6 yields the following:

$$\frac{dq}{dt} = \xi(P_1 + [L - 1]q) + \gamma q(1 - q)\{b(1 - \alpha) + \alpha(2q - 1)\} \quad (11)$$

Using this expression we can follow the same procedure as before to derive the conditions for *S*-curves. Note the similarity between Equations 14 and 9:

$$b > \frac{\xi(1 - L) + \gamma\alpha}{\gamma(1 - \alpha)} = \frac{\phi(1 - L) + \alpha}{(1 - \alpha)} \quad (12)$$

Figure 10 illustrates the curves for Equation 12 when  $\phi = 1$ . As the strength of conformist transmission increases, the region of  $b$  and  $L$  values that generates *S*-curves shrinks—remember, *S*-curves begin appearing as one moves to the right of, or above, the curve. Consequently, if conformist transmission is even a small component of human psychology, we should expect either (1) that all the various values of  $b$  represented throughout the diffusion of innovations literature are quite high or (2) that the value of  $L$  (or  $d$ ) in human psychology is substantial—otherwise *R*-curves would be more common.

Returning to the question I asked at the beginning of this section: Can conformist transmission account for the long tail observed in Figure 1? Figure 11 shows the temporal dynamics for a series of  $\alpha$  values, ranging from 0 to 0.27. Comparing Figure 1 and Figure 11, we observe that conformist transmission does generate the long tails observed in some empirical data. Assuming that  $\alpha$  is fairly small, such tails occur when the biases generated by the nonconformist components of our cultural capacities are relatively weak ( $b \approx \alpha$ ). When these biases are large ( $b > \alpha$ ), the effect of  $\alpha$  disappears.

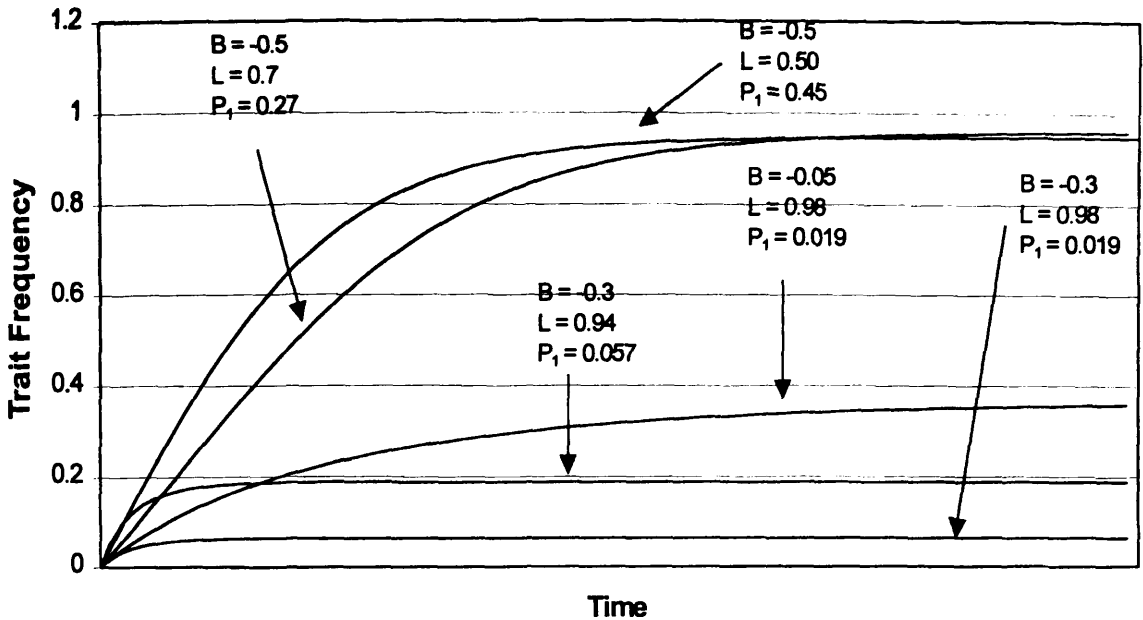


Figure 8. Temporal dynamics when ENLR spreads a novel trait against BCT for five sets of parameters. None of these is an S-curve.

More generally, this slow growth period is a common feature of many adoption curves. Rogers (1995:259–260) explains that potential adopters initially seem resistant to new ideas until a “critical mass” is achieved and the diffusional process “takes off.” This intuitive explanation supports the idea, formalized in conformist transmission (Boyd and Richerson 1985), that individuals use the frequency of a trait as an indirect indicator of its worth. Hence, a trait’s frequency inhibits its diffusion when it is rare but encourages the diffusion once the trait becomes common.

Conformist transmission can also help in predicting the takeoff points described by applied diffusion researchers. In attempting to actively spread novel innovations, governments, states, and organizations will sometimes provide “pump-priming” incentives to adopters, often in the form of direct cash payments, until the innovation spreads past some critical frequency (often thought to lie between 20 and 30 percent of potential adopters; Rogers 1995:221). Once this threshold is reached, the innovation is considered self-sustaining, which means that it will continue to spread on its own. Note that the empirical existence of takeoff points supports both the claim that  $L$  is big and the claim that conformist transmission is real (but small). If we do not assume that  $L$  is big, then environmental learning will always spread beneficial traits and takeoff points should not exist. Assuming that  $L$  is big, we can derive a simple expression for the takeoff frequency using Equation

10.<sup>9</sup> With this assumption, the diffusional process becomes self-sustaining when

$$B = b(1 - \alpha) + \alpha(2q - 1) > 0 \quad (13)$$

A diffusional process is not self-sustaining when the magnitude of the conformist component of  $B$ ,  $|\alpha(2q - 1)|$ , exceeds  $b(1 - \alpha)$ , thereby making the overall bias less than 0. Remember, the conformist component is negative when  $q < 0.50$ . Solving Equation 13 for the takeoff frequency,  $q_{\text{tp}}$ , requires setting  $B = 0$  and solving for  $q$ :

$$q_{\text{tp}} = \frac{1}{2} - \frac{b(1 - \alpha)}{2\alpha} \quad (14)$$

This is the value of  $q$  at the point where  $B$  crosses over from negative to positive values.

Equation 14 tells us two things. First, if it exists, the takeoff frequency lies between 0 and 0.50. And second, if  $b(1 - \alpha)/2\alpha > 0.5$ , then the process will never be self-sustaining. Empirical data indicate that pump-priming incentives do often work (but not always) and that takeoff points always seem to lie between 0 and 0.5 (Rogers 1995).<sup>10</sup>

### Modifying the Environmental Learning Model Still Will Not Produce S-Shapes

A great deal of research on diffusion has adopted the intuition that diffusional dynamics, including the S-shape, result from differences among individuals in their degree of “innovativeness” or their fear of uncertainty. For example, Rogers writes:

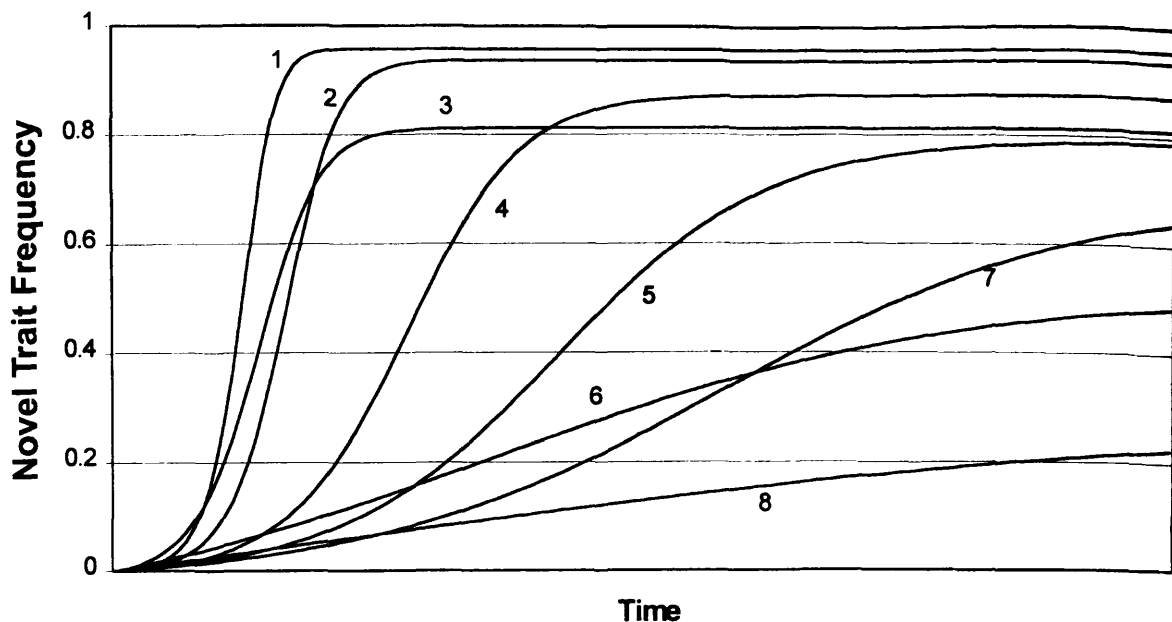


Figure 9. Temporal dynamics when biased transmission spreads a novel trait against individual learning for eight sets of parameters. Values of  $B$ ,  $L$ , and  $P_i$  (respectively) for Curve 1: 0.45, 0.98, 0.001; Curve 2: 0.30, 0.98, 0.001; Curve 3: 0.30, 0.94, 0.003; Curve 4: 0.15, 0.98, 0.001; Curve 5: 0.90, 0.80, 0.01; Curve 6: 0.5, 0.7, 0.03; Curve 7: 0.6, 0.8, 0.01; and Curve 8: 0.2, 0.8, 0.02. For Curve 8,  $B = \phi(1 - L)$ , so this is technically an  $R$ -curve, but its low equilibrium frequency means that it probably would not "count" as a successful diffusion ( $q_{eq} = 0.32$ ).

Many human traits are normally distributed, whether the trait is a physical characteristic, such as weight or height, or a behavioral trait, such as intelligence or the learning of information. Hence, a variable such as the degree of innovativeness is expected also to be normally distributed. [1995:258]

The idea is that a few individuals with a high degree of innovativeness adopt early, most people adopt somewhere in the middle, and a few stragglers, with low innovativeness, adopt late. Although it may be true that individuals vary in their degree of innovativeness,<sup>11</sup> building this into the environmental learning model does not produce the anticipated  $S$ -dynamics, as I will demonstrate. Furthermore, I have already shown that  $S$ -dynamics can be produced without assuming that people are different (see also Cavalli-Sforza and Feldman 1981). All the models so far have assumed that people are psychologically and socially identical, yet they still produce  $S$ -curves under a wide range of conditions.

We can construct environmental learning models that incorporate individual variation in two ways: (1) assume individuals do environmental learning first and then, if they remain uncertain, rely on *unbiased* transmission (copy someone at random)—this provides a transgenerational model (guided variation); or (2) assume that individuals do repeated trials and that the dynamics of learning are fast relative to an individual's lifetime (or that individuals live

forever). In the previously described environmental learning model, the parameter  $d$  (see Figure 2) represents an individual's threshold of evidence or his or her willingness to proceed under uncertainty. Innovative individuals are those willing to adopt a new trait based on limited (uncertain) evidence. Thus, this parameter captures what many researchers mean by "innovativeness." Following the standard approach from the diffusion literature—to classify people into adopter categories—I define five types of individuals: innovators, early adopters, early majority, later majority, and laggards (Rogers 1995:262). The subscript  $i$  indexes these categories from 1 to  $n$  ( $n = 5$  in this case). Each category  $i$  is characterized by its own value,  $d_i$ . Innovators have the smallest value of  $d$ , and laggards have the largest value of  $d$ . Each value of  $d_i$  generates, via the cumulative normal distribution shown in Figure 2, corresponding values of  $P_{ii}$  and  $L_i$ . For the first version of the model (with unbiased transmission), the frequency of the novel trait among members of category  $i$  (e.g., early adopters) in the next time cycle is shown in Equation 15:

$$q'_i = P_{ii} + qL_i \quad (15)$$

Further, assume that the symbol  $F_i$  represents the proportion of the total population that adopter category  $i$  comprises. For example, if 10 percent of the population are laggards, then  $F_5 = 0.10$ . To find the new frequency of the

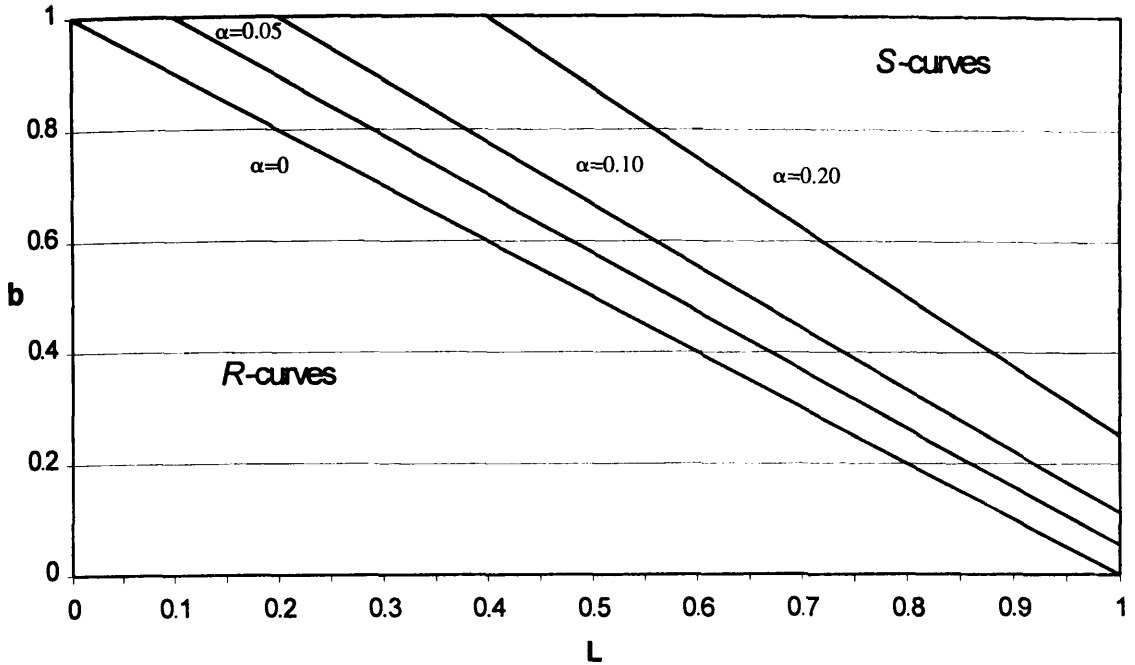


Figure 10. S-curve regions for different strengths of conformist transmission ( $\phi = 1$ ). Increasing the strength of conformist transmission,  $\alpha$ , reduces the size of the B-L region that produces S-curves. However, given that other evidence indicates the existence of a conformist component, this plot further indicates that  $L$  (or  $d$ ) is probably large.

novel trait in the overall population, we compute the expected value of Equation 15:

$$q' = E(q'_i) = \sum_i F_i(P_{1i}) + \sum_i F_i(L_i, q) = \bar{P}_1 + q\bar{L} \tag{16}$$

Equation 16 demonstrates that, when individuals vary in their innovativeness, the cumulative adoption curves depend only on the average values of  $P_1$  and  $L$  (regardless of their distribution). This means that Equation 16 behaves just like Equation 1 and, therefore, does not produce S-curves.

In the second version of individual variation, instead of unbiased transmission, I assume that the dynamics of learning are fast relative to the lifetimes of individuals. Thus, we get Equation 17:

$$q'_i = P_{1i} + qL_i \tag{17}$$

The only difference between Equations 17 and 15 is the subscript  $i$  on  $q$ . This occurs because, instead of copying someone at random from the population every time period, these long-lived individuals simply stick with their current behavior—that is, the environment does not provide sufficiently convincing data to justify a change. Taking the expectation of  $q'_i$  to get  $q'$ , we arrive at Equation 18:

$$q' = E(q'_i) = E(P_{1i}) + E(L_i, q) = \bar{P}_1 + q\bar{L} + COV(q_i, L_i) \tag{18}$$

NS

Does Equation 18 produce S-curves? We have already seen that the terms labeled “NS” in Equation 18 will never produce an S-curve. The final term,  $COV(q_i, L_i)$ , is the covariation between  $q_i$  and  $L_i$ , which varies for different values of  $q$  (or over time). At  $t = 0$  (and  $q = 0$ ),  $COV = 0$ . In Figure 12, although the frequency of Trait 1 rises for each of the subgroups in the population ( $q_i$ ), the subgroup with the smallest value of  $d_i$  learns the novel trait most quickly—note that the different values of  $q_i$  can be observed at the points where the vertical line crosses the different curves (which have different values of  $d_i$ ). The initially rapid adoption of the trait by more innovative individuals (those with lower  $d$  values) generates a negative covariation between  $L_i$  and  $q_i$ . This negative association remains until the curves cross one another in the middle of Figure 12. After this crossover, the  $COV(q_i, L_i)$  crosses through 0 and stabilizes at a positive equilibrium value. These dynamics for covariation remain robust because more innovative individuals adopt novel behaviors more rapidly but achieve lower equilibrium values of  $q_i$  than less innovative individuals. Lower equilibrium values

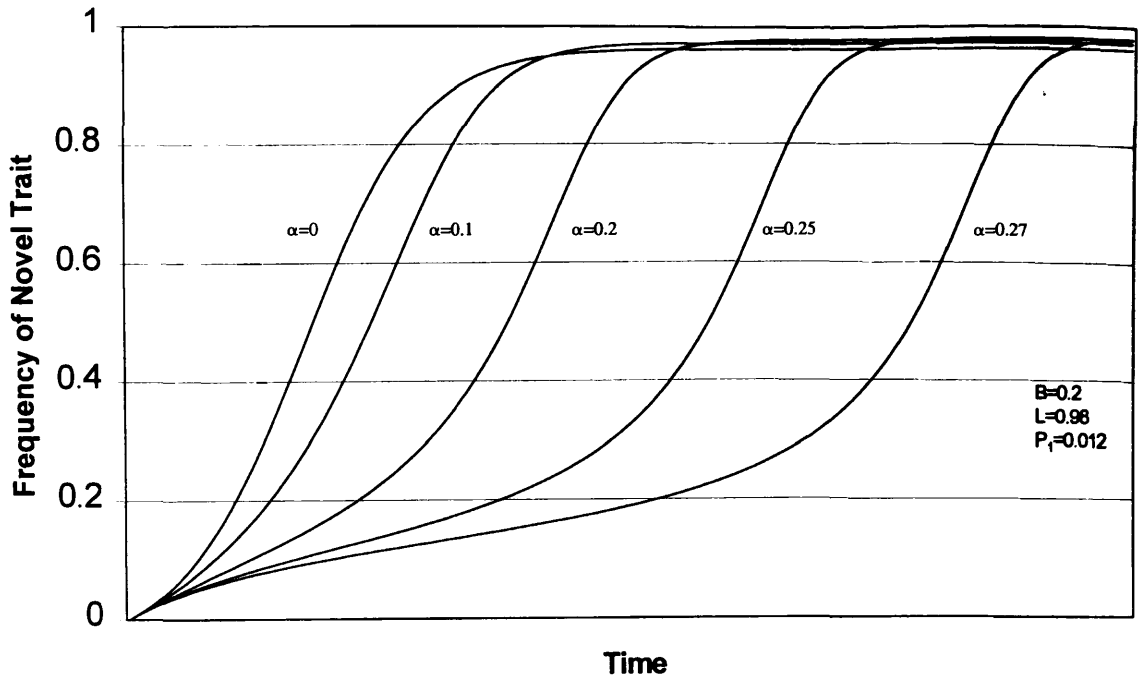


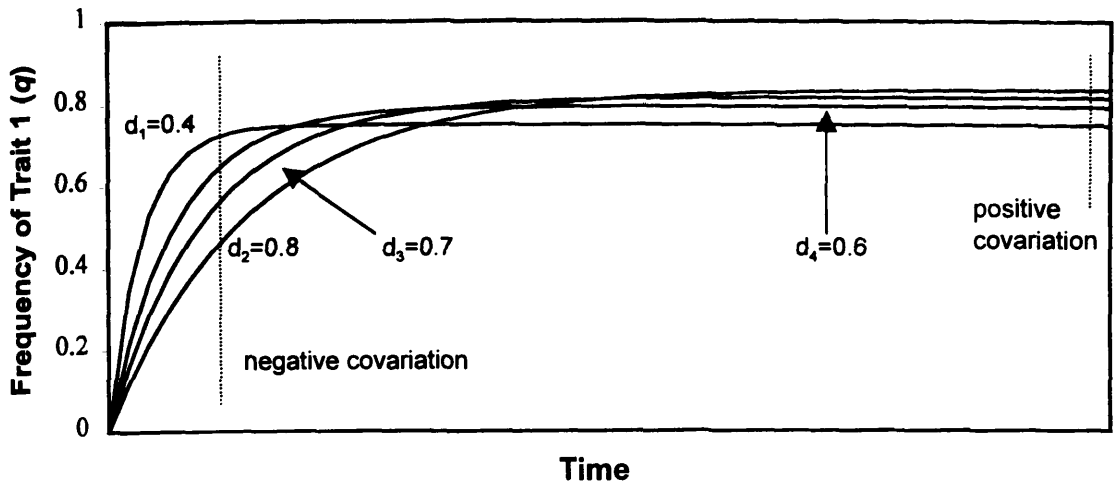
Figure 11. Adoption curves for different values of  $\alpha$ . As the size of the conformist component increases, the initial takeoff stretches out. Eventually, as  $\alpha$  rises, the diffusion of the novel trait will be prevented by conformist transmission.

occur because more innovative individuals are subject to more erroneous switchbacks, as their standard of evidence for changing behaviors is lower—it is the price of innovativeness.<sup>12</sup> The dynamics of  $COV(q, L_t)$ , when added to the standard *R*-shaped curves produced by the NS terms in Equation 18, will never produce an *S*-curve.

Many researchers believe the *S*-shaped cumulative adoption curve to be the product of an underlying, normally distributed “time-to-adoption” curve that captures the varying degrees of innovativeness distributed throughout the population. In this view, “time to adoption” acts as the inverse of innovativeness. When researchers test their empirically derived  $\Delta q/\Delta t$  ( $dq/dt$ ) curves for deviations from normality, sometimes they pass (and cannot be distinguished from a normal distribution), and sometimes they do not. When such curves do not pass the normality test, researchers claim that they “approach normality.” For example, using the Iowa farmer data (Figure 1), researchers went to great lengths to show that the data were normally distributed. Yet they failed to show normality because of the distribution’s long tail. However, from the perspective I have presented here, there is no reason to expect underlying normality. Often, Equation 11 does produce time-to-adoption distributions that look approximately normal, but knowing if they are approximately normal or not does not tell us anything more about the un-

derlying social-decision processes. For example, the time derivative of a logistic curve (which is appropriate for biased transmission) looks quite normal and would certainly appear normal if one sampled from it. More importantly, Equation 11 can also produce underlying, nonnormal, time-to-adoption distributions that are much more similar to those produced by the diffusion of hybrid corn or of Tetracycline than to any normal distribution.

Many efforts to fit the *S*-dynamics of the diffusion literature have been made, especially in the marketing and new product literatures (e.g., Bass 1969; Jensen 1982; Kalish 1985; Oren and Schwartz 1988). For a long time, researchers have recognized that logistic curves in various forms can fit many of the *S*-curves fairly well. Unfortunately, the parameters in these functional forms have little meaning because such “curve fits” lack any a priori theoretical foundation in human psychology or decision making (Bass 1969). However, by assuming that individuals vary in their degree of risk aversion in particular ways, some researchers have managed to construct environmental learning models that under some conditions will generate logistic *S*-curves (Jensen 1982; Kalish 1985; Oren and Schwartz 1988). Although these models can produce *S*-curves, based on individual differences in risk aversion and Bayesian learning processes, the circumstances that produce the *S*-dynamics depend critically on the initial



**Figure 12.** Environmental learning dynamics for four subgroups with differing degrees of innovativeness ( $d_i$ ). During the first portion of the diffusion a negative covariation builds up between  $d_i$  and  $q_i$ , but then as certain categories of people reach equilibrium the covariation crosses over and becomes positive. This occurs because of the aforementioned tradeoff between speed and accuracy in environmental learning.

distribution of beliefs in the populations, the specific shape of the utility curves, and the details of the information-gathering processes. In Oren and Schwartz's model (1988), for example, deriving the logistic form depends on assuming both a constant proportional risk aversion ( $xU''[x]/U'[x] = \text{constant}$ ) and that risk aversion is *exponentially* distributed across the population. No empirical justification for either of these rather narrow assumptions is provided. Without any empirical support, it is difficult to believe that these assumptions are as robust across the world's populations as are the  $S$ -dynamics of diffusion. Similarly, under some conditions the environmental learning model in this article will produce  $S$ -curves if innovative individuals (those with low thresholds of evidence) are assumed to acquire or process information better than less innovative people. However, getting an  $S$  depends on exactly how innovativeness and information-processing abilities are distributed across the population.

Some readers may criticize this analysis because they realize that a wide variety of mathematical formulations of environmental learning or rational calculations could generate  $S$ -curves, and I have not begun to exhaust the possible formulations. This is true. However, merely having equations with the symbols arranged in a particular fashion is not a sufficient riposte. In my view, the trick is to formulate a learning model that is rooted in human psychology, is evolutionarily plausible, is empirically grounded in what we know about human cognition, and still produces  $S$ -curves under a wide range of general conditions. I hope that skeptics who favor environmental learning will endeavor to generate and test such competing models.

## Discussion and Summary

Many scholars have the intuition that cultural transmission is, at best, a minor force in human behavior and behavioral change (Buss 1999; Harris 1979; Pinker 1997; Stigler and Becker 1977; Tooby and Cosmides 1992). However, if cultural transmission is merely a weak component of the psychological processes that generate human behavior—meaning  $L$  is fairly small—then we would expect the real world, and the diffusion of innovations literature, to contain a large proportion of  $R$ -curves relative to the proportion of  $S$ -curves. If people have small  $L$  values,  $S$ -curves should result only when the replicatory bias ( $B$ ) is quite high. However, as I mentioned earlier,  $B$  is generated in part by the qualities of the trait itself (e.g., eating high-fat foods or believing in a good god) and by the qualities of the trait's possessors (i.e., their local prestige or success). So  $B$  values should vary substantially from situation to situation, and (if  $L$  is small) we should observe  $R$ -curves when  $B$  is medium or low. However, in the real world (or at least in the available empirical data),  $R$ -curves are relatively rare while  $S$ -curves are rampant. This suggests that biased cultural transmission dominates the diffusional process and that  $L$  must be pretty big—or somehow hundreds of researchers studying everything from the spread of insecticides among Colombian peasants to the diffusion of “poison pills” among Fortune 500 companies must have systematically biased the database and selected only traits with very high bias values ( $B$ ).

Further evidence for a substantial reliance on cultural transmission comes from the spread of maladaptive or costly behavioral traits. My analysis indicates that



maladaptive traits may spread against the force of individual learning—to produce an *S*-curve—as long as *L* and *B* are sufficiently large. For example, the practice of bottle-feeding infants spread throughout the Third World despite the fact that this practice produces higher rates of sickness, infection, and death in infants under Third World conditions than does breast-feeding (Rogers 1995). Such costly, maladaptive practices abound in the anthropological literature (Edgerton 1992). In many societies, food taboos restrict the consumption of nutritionally valuable foods (Baksh 1984; Descola 1994; Wilbert 1993). Even in places where protein and dietary fat are limited, people still refuse to eat valuable nutritional resources. The Machiguenga of the Peruvian Amazon, for example, would not consider eating snake meat, even when the dead snake is known to be nonvenomous. Similarly, the Warao, who inhabit the extremely marginal environs of the Orinoco River delta, refuse to hunt large mammals (which include some of the most valuable animal resources in South America) because they “have blood like people” (Wilbert 1993:18). Furthermore, nearly one-half of all cultures surveyed throw out the valuable colostrum that precedes mother’s milk and helps infants develop their immune systems while providing essential minerals (Morse et al. 1990). Without the predominance of biased transmission, it would be difficult to explain the prevalence of costly, maladaptive traits in populations throughout the world. Remember, most environmental learning models, like Oren and Schwartz’s (1988), predict that only beneficial, utility-maximizing, or adaptive traits will spread through populations, but that is not all we observe. Finally, even if we assume that different “cultural logics” could cause individuals from particular groups to evaluate the costs and benefits of seemingly “maladaptive practices” (“maladaptive” to an outsider) in such a way as to favor their spread, we would still expect to observe *R*-shaped diffusion curves (if environmental learning is the dominant process). Yet we do not.

On the flip side, if our reliance on biased transmission were weak (if *L* were small), then environmental learning would frequently spread beneficial traits against the tide of negatively biased cultural transmission. However, my analysis indicates that we should record an *R*-curve every time our cost-benefit analysis overcomes our social learning tendencies. Yet *R*-curves are rare, so biased transmission is most likely a substantial component of human behavioral plasticity.

Finally, how can environmental, cost-benefit learning account for the empirical phenomena of long tails and takeoff points? Why do diffusional processes sometimes begin so slowly and finish so rapidly? Why doesn’t this occur other times? Why do some behaviors have threshold adoption frequencies at which they begin spreading on their own (without people being paid for adoption), even when the behavior later turns out to be a bad idea? As I have described, the simple models of biased transmission

(with a small conformist component) presented in this article can account for all these phenomena, but it appears unlikely that the environmental learning approach, based on the direct evaluation of payoffs, can be modified to account for them as well.

### What Kind of Information Flows through Social Networks?

Many social scientists believe that by diffusing “information,” social networks generate the classical diffusional dynamics. Rogers writes that “[diffusional networks] convey information to decrease uncertainty about a new idea” (1995:281). By using the term *innovation-evaluation information* Rogers captures what I describe above as “payoff-relevant” information, which is the essential ingredient in the individual-level environmental learning model (and most cost-benefit approaches). Although the biased cultural transmission processes I have modeled here do involve the transfer of information among individuals, this imitation process does *not* directly involve the transmission of innovation-evaluation information—that is, information used by individuals to evaluate the costs and benefits of alternative practices. Biased imitation involves copying an idea or practice for reasons not directly related to its costs and benefits. From this perspective, information remains the key element flowing through social networks, but rather than information about payoffs (direct cost-benefit information), it is information about such things as who have adopted a particular practice (how prestigious they are) or how many others have adopted the practices. The available empirical data (from both the field and the laboratory) support the kinds of imitation processes I have described, not the innovation-evaluation hypothesis.

For example, in prestige-biased transmission, individuals copy traits possessed by prestigious individuals, regardless of whether these traits affect the success of the prestigious model or the copier (Henrich and Gil-White 2001 summarizes both theory and evidence). Generally, the enormous importance of what diffusion researchers call “opinion leadership” confirms the theoretical predictions of prestige-biased transmission (Rogers 1995:293). For example, the same farming practice will spread rapidly in places where the local high-prestige individuals favor the novel idea but will entirely fail to spread in other places where the prestigious individuals reject the novel practice. Similarly, Van den Ban (quoted in Rogers 1995) effectively demonstrates the importance of prestige-biased transmission over evaluative information processing in his study of farmers in the Netherlands. He shows that small-scale farmers copied the farming practices of prestigious, large-scale farmers *even* when such practices were clearly inappropriate for their particular situation.

Like prestige-biased transmission, conformist transmission does not depend *directly* on the costs and benefits of

alternative behaviors, but it still seems to be an important component of adoption dynamics. Besides the long tails and takeoff points observed in many diffusional curves, conformist transmission can also account for the spatial or sociospatial clustering of traits frequently observed in the diffusion literature. For example, in studying the spread of contraceptive methods in rural Korean villages, Rogers and Kincaid (1981) found that choices clustered by village: there were "pill villages," "IUD villages," and even "vasectomy villages." All these contraceptive methods were being promoted equally by the government campaign, and each village contained individuals with differing degrees of wealth and social standing. Cost-benefit analyses, environmental learning, and most kinds of direct biases can neither generate nor maintain such patterns. Eventually, given any social connection among villages (which Rogers and Kincaid did clearly observe), the contraceptive method with the highest bias or greatest benefits ratio should spread to all villages. Or if all methods were somehow exactly equal in benefits and costs or direct biases, then we would expect these methods to scatter across the social landscape and not cluster in village networks. In contrast, conformist transmission predicts sociospatial clusters of similar traits any time the differences between the costs and benefits or the biases of alternative practices are relatively small. Similar patterns of innovation clusters were observed by Whyte (1954) in his study of the spread of air-conditioning units in Philadelphia.<sup>13</sup>

### Cultural Drift Can Produce S-Shaped Diffusions

The models in this article have ignored the effects of random sampling errors in transmission on the diffusional process by assuming that the forces of environmental learning and biased transmission are large compared with the effects of sampling errors. However, this sampling error, or "cultural drift," may be important if both  $\mu$  (the expected perceivable payoff difference) and  $B$  are quite small. With  $B$  close to 0, individuals can be thought of as relying on unbiased transmission. Because any particular sample of cultural models from the population will not exactly represent the current distribution of traits in the population, repeated sampling through successive time steps may sometimes accumulate successive "errors" favoring one trait or another. Under these conditions, particularly in smaller groups, a trait may occasionally be driven by repeated sampling errors to high frequency and fixation. Such drift-driven diffusions will be S-shaped.

Although drift-driven processes may be responsible for some S-shaped diffusions, this process seems unlikely to account for most of the diffusions observed in the innovations literature. I have three reasons for this. First, drift-driven diffusions will occur suddenly without any correspondence to when a particular trait first entered a population. In contrast, many of the diffusions documented

in the innovations literature occur directly on the heels of the introduction of a novel practice, product, idea, or technology to a particular social group. Consistent with this, learning processes should begin affecting the frequency of a novel trait soon after its introduction. Cultural drift, however, is no more likely to begin driving an adoption immediately after an introduction than it is 100 years later. Second, drift-driven diffusions will usually occur rather rapidly (Hartl and Clark 1989), so the S-curves will lack the "long tail" that characterizes many diffusions. In addition, the conformist component of cultural transmission, which acts against any trait at low frequency, means that drift-driven diffusions should be even more rare than we would expect under unbiased transmission—although when such diffusions do occur they will often have "long heads" because once cultural drift drives (usually quite rapidly) a trait's frequency above 50 percent, conformist transmission can drive it gradually to fixation. Such curves will have a slow finish, rather than a slow start. And, third, drift-driven diffusions should occur *less* frequently in larger populations than in smaller populations because sampling errors are more likely to accumulate in small groups (Cavalli-Sforza and Feldman 1981). Yet larger populations seem to experience more frequent diffusions than smaller, more isolated social groups—a fact often attributed to economic differences and risk aversion (Henrich and McElreath 2000). However, unlike the case for areas typically studied by diffusion researchers, it seems plausible that some of the diffusions observed in the archaeological record—stylistic variations, for example—could result from cultural drift (Lipo et al. 1995; Neiman 1995).

A great deal of empirical work has been done on the characteristics of "innovators" and "early adopters"—those who adopt early in diffusional processes (Cancian 1979, 1989; Dewees and Hawkes 1988; Rogers 1962, 1995). At first glance, these patterns are convincing. According to this work, early adopters tend to have larger social networks, higher status, more money, more cosmopolitan contacts, and more exposure to mass media outlets. The assumption seems to be that these characteristics (causally) increase an individual's likelihood of adopting an innovation early in the diffusional process. Unfortunately, the literature's focus on successful diffusions produces an extremely biased database. The only situations included in the database involve those in which the trait actually spread; in contrast, all those times when the trait did not spread are *not* included. So, the more accurate empirical claim would be: early adopters tend to have larger networks, higher status, and so on *given that the trait eventually spreads to high frequency*. It is quite possible that all individuals, regardless of their economic positions, media exposure, and so forth, are *equally* likely to adopt an innovation early but that the *subsequent* diffusion of an

"innovation" depends on the characteristics of the initial adopters. Things such as large social networks and high status may have nothing to do with an individual's chances of innovating, but they may be critical to the subsequent transmission of these traits. When poor, low-status individuals innovate, nobody copies them; so the trait never diffuses, and the individuals never get into the database as "innovators."

Future work should turn the problem of diffusion on its head and explain why certain societies, particularly peasant groups, seem slow or resistant to the spread of novel behavioral traits, ideas, and "innovations." Building on theoretical work that suggests that individuals should increase their  $L$  and  $\alpha$  values as environmental information becomes ambiguous or problems become difficult (Henrich and Boyd 1998), peasant researchers might explore the connections among information quality (e.g., yield variance), social learning, conformist transmission, and rates of adoption. As well, this work could address how cultural transmission mechanisms, under certain circumstances, can produce upper-middle-class conservatism or the "Cancian-dip" (Cancian 1979). Finally, such work could use diffusional data from a wide variety of sources and numerical computer simulations to estimate parameter distributions for  $L$ ,  $P$ ,  $b$ , and  $\alpha$ .

## Notes

*Acknowledgments.* I would like to thank Natalie Smith, Robert Boyd, Allen Johnson, Nick Blurton Jones, Peter Richerson, R. Lee Lyman, and the four anonymous reviewers for their helpful comments and suggestions.

1. Recently, a vast amount of work in cognitive psychology and experimental economics has severely criticized the extreme, hyper-rational models of classical economics (e.g., Gigerenzer and Goldstein 1996; Henrich 2002; Kagel and Roth 1995; Kahneman et al. 1982; Rabin 1998); consequently, many economists and other students of human behavior are increasingly turning to cognitively more realistic models of human learning and decision making.

2. Throughout this article, I use *behavioral trait*, or simply *trait*, to stand for a whole range of things that could be "innovations," "cultural traits," "practices," "beliefs," "ideas," or "values."

3. This kind of experimentation is common in both traditional and modern agricultural systems—see Chibnik 1981, Johnson 1972, and Rogers 1995.

4. By "substantially," I mean an effect of the same order of magnitude or larger than the cost-benefit effect.

5. I think some of the confusion (especially among archaeologists) about direct bias arises from reading Bettinger's description (1991:188). This description is not wrong, but it is incomplete and rather misleading in the context of this article. At its most general level, direct bias can arise from anything about the interaction between human psychologies (e.g., preferences, ease of storage or recall from memory, goals) and particular cultural traits that make individuals more likely to

acquire one cultural variant over others (also see Boyd and Richerson 1985:135). Note that, in general, I find that Bettinger (1991) provides an excellent, insightful summary of cultural transmission theory.

6. Prestige-biased transmission represents a particular type of indirect bias (Bettinger 1991:196; Boyd and Richerson 1985: ch. 8) for which we have a substantial amount of empirical evidence (Henrich and Gil-White 2001).

7. Prestige biases may be either constant or frequency dependent (dependent on  $q$ ), depending on whether the frequency of the transmitted trait in the population significantly affects the success or prestige of the trait's possessors. This article does not incorporate frequency-dependent biases, except for conformist transmission, but many transmission models have built this in using evolutionary game theory (e.g., McElreath et al. 2000).

8. This occurs because whichever equation is second in the life cycle (meaning whichever one gets substituted into) exerts a small bias on the final result. It is a sampling bias that favors the most recent recursion.

9. Furthermore, if we do not assume that  $L$  is big, then Equation 10 yields a cubic equation in  $q$ , which can be solved but does not yield any useful insights.

10. In analyses not discussed in this article, I have added conformist transmission to the situations in which biased transmission and environmental learning oppose one another. I have found that this modification does not qualitatively change the results discussed in these sections.

11. Looking across several domains of "innovation," Dewees and Hawkes (1988) found that particular commercial fishermen could not be generally characterized as "innovators" or "laggards" in their study of six different fishing-related innovations. Individuals who adopted early in one instance might not adopt at all in another instance. "Innovativeness" may not be a dispositional trait valid across many behavioral domains.

12. The mean of the normal distribution shown in Figure 2 represents the quality of environmental information that is available to every individual—this variable tells us how difficult the problem is. In this model, I assume that everyone receives the same quality of information and has the same abilities to process this information.

13. There are other explanations for this kind of clustering besides conformist transmission, including combinations of other types of cultural transmission mechanisms (see Henrich and Boyd 1998). Another possibility is that if the costs or benefits of an innovation were frequency dependent, then once one method attains high frequency, by whatever stochastic processes, it would remain at high frequency. In some situations, this hypothesis seems mildly plausible, but in other situations, like the spread of contraceptive methods, it is difficult to see the frequency dependence.

## References Cited

- Baksh, Michael G.  
1984 Cultural Ecology and Change of the Machiguenga Indians of the Peruvian Amazon. Ph.D. dissertation, University of California at Los Angeles.

- Bandura, Albert**  
1977 *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Baron, R., J. Vandello, and B. Brunsman**  
1996 The Forgotten Variable in Conformity Research: Impact of Task Importance on Social Influence. *Journal of Personality and Social Psychology* 71(5):915–927.
- Bass, Frank M.**  
1969 A New Product Growth Model for Consumer Durables. *Management Science* 15:215–227.
- Bettinger, Robert L.**  
1991 *Hunter-Gatherers: Archaeological and Evolutionary Theory*. New York: Plenum Press.
- Bettinger, Robert L., Robert Boyd, and Peter Richerson**  
1996 Style, Function and Cultural Evolutionary Processes. *In Darwinian Archaeologies*. Herbert D. G. Maschner, ed. Pp. 133–164. New York: Plenum Press.
- Binford, Lewis**  
1983 *Working at Archaeology*. New York: Academic Press.
- Bowles, Samuel**  
1998 Cultural Group Selection and Human Social Structure: The Effects of Segmentation, Egalitarianism, and Conformism. Electronic document, <http://www-unix.oit.umass.edu/bowles/papers/cultural.PDF>.
- Boyd, Robert, and Peter J. Richerson**  
1985 *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.  
1988 An Evolutionary Model of Social Learning: The Effects of Spatial and Temporal Variation. *In Social Learning: Psychological and Biological Perspectives*. Thomas R. Zentall and Bennett G. Galef, eds. Pp. 29–48. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Buss, David**  
1999 *Evolutionary Psychology: The New Science of the Mind*. Boston: Allyn and Bacon.
- Camerer, Colin, and Teck-Hua Ho**  
2000 Experience-Weighted Attraction Learning in Normal Form Games. *Econometrica* 67(4):827–875.
- Cancian, Frank**  
1979 *The Innovator's Situation: Upper Middle Class Conservatism in Agricultural Communities*. Stanford: Stanford University Press.  
1989 Economic Behavior in Peasant Communities. *In Economic Anthropology*. Stuart Plattner, ed. Pp. 127–170. Stanford: Stanford University Press.
- Cavalli-Sforza, Luca L., and Marcus Feldman**  
1981 *Cultural Transmission and Evolution*. Princeton: Princeton University Press.
- Chibnik, Michael**  
1981 The Evolution of Cultural Rules. *Journal of Anthropological Research* 37(3):256–268.
- Coleman, James S., Elihu Katz, and Herbert Menzel**  
1966 *Medical Innovation: A Diffusion Study*. New York: Bobbs-Merrill.
- Descola, Philippe**  
1994 *In the Society of Nature: A Native Ecology in Amazonia*. New York: Cambridge University Press.
- Deweese, C. M., and G. R. Hawkes**  
1988 Technical Innovation in the Pacific Coast Trawl Fishery: The Effects of Fishermen's Characteristics and Perceptions on Adoption Behavior. *Human Organization* 47(3):224–234.
- Durham, William H.**  
1991 *Coevolution: Genes, Culture, and Human Diversity*. Stanford: Stanford University Press.  
1992 *Applications of Evolutionary Culture Theory*. *Annual Review of Anthropology* 21:331–355.
- Earle, Timothy**  
1997 *How Chiefs Come to Power*. Stanford: Stanford University Press.
- Edgerton, Robert B.**  
1992 *Sick Societies: Challenging the Myth of Primitive Harmony*. New York: Free Press.
- Erev, Ido, and Alvin Roth**  
2000 Learning, Reciprocation and the Value of Bounded Rationality. *In Bounded Rationality: The Adaptive Toolbox*. Gerd Gigerenzer and Reinhard Selten, eds. Pp. 215–231. Cambridge: MIT Press.
- Gigerenzer, G., and D. G. Goldstein**  
1996 Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review* 103(4):650–669.
- Gintis, Herbert**  
2000 *Game Theory Evolving*. Princeton: Princeton University Press.
- Gladwin, C., and J. Butler**  
1984 Is Gardening an Adaptive Strategy for Florida Family Farmers? *Human Organization* 43(3):208–216.
- Harris, Marvin**  
1979 *Cultural Materialism: The Struggle for a Science of Culture*. New York: Random House.
- Hartl, Daniel L., and Andrew G. Clark**  
1989 *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates.
- Henrich, Joseph**  
2002 Decision-Making, Cultural Transmission and Adaptation in Economic Anthropology. *In Economic Anthropology: Theory at the Turn of the Century*. J. Ensminger, ed. New York: Rowman and Littlefield Press. Forthcoming.
- Henrich, Joseph, and Robert Boyd**  
1998 The Evolution of Conformist Transmission and the Emergence of Between-Group Differences. *Evolution and Human Behavior* 19:215–242.  
2001 On Modeling Cognition and Culture: How Formal Models of Social Learning Can Inform Our Understanding of Cultural Evolution. Electronic document, <http://webuser.bus.umich.edu/henrich/>.
- Henrich, Joseph, and F. Gil-White**  
2001 The Evolution of Prestige. *Evolution and Human Behavior* 22(3):165–196.
- Henrich, Joseph, and Richerson McElreath**  
2000 Are Peasants Risk Averse Decision-Makers? Electronic document, <http://webuser.bus.umich.edu/henrich/risk.pdf>.
- Insko, C. A., R. H. Smith, M. D. Alicke, J. Wade, and S. Taylor**  
1985 Conformity and Group Size: The Concern with Being Right and the Concern with Being Liked. *Personality and Social Psychology Bulletin* 11(1):41–50.

- Jensen, Richard  
1982 Adoption and Diffusion of an Innovation of Uncertain Profitability. *Journal of Economic Theory* 27:182–193.
- Johnson, Allen  
1972 Individuality and Experimentation in Traditional Agriculture. *Human Ecology* 1(2):149–159.
- Jovanovic, B., and Y. Nyarko  
1995 A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves. *Brookings Papers on Economic Activity* 1:247–299.
- Kagel, John H., and Alvin E. Roth  
1995 *The Handbook of Experimental Economics*. Princeton: Princeton University Press.
- Kahneman, Daniel, Paul Slovic, and Amos Tversky  
1982 *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Kalish, Shlome  
1985 New Product Diffusion Model with Price, Advertising and Uncertainty. *Management Science* 31:569–585.
- Kawai, Masao  
1965 Newly Acquired Pre-Cultural Behavior of the Natural Troop of Japanese Macaques. *Primates* 6:1–30.
- Lefebvre, Louis, and Luc-Alain Giraldeau  
1994 Cultural Transmission in Pigeons Is Affected by the Number of Tutors and Bystanders Present. *Animal Behaviour* 47(2):331–337.
- Lipo, C., M. Madsen, R. C. Dunnell, and T. Hunt  
1995 Population Structure, Cultural Transmission, and Frequency Seriation. *Journal of Anthropological Archaeology* 16:301–334.
- McElreath, Richard, Robert Boyd, and Peter Richerson  
2000 Shared Norms Can Lead to the Evolution of Ethnic Markers. Electronic document, <http://www.sscnet.ucla.edu/anthro/faculty/boyd/EthnicMarkers1.8.pdf>.
- Morse, J. M., C. Jehle, and D. Gamble  
1990 Initiating Breastfeeding: A World Survey of the Timing of Postpartum Breastfeeding. *International Journal of Nursing Studies* 27(3):303–313.
- Neiman, Fraser D.  
1995 Stylistic Variation in Evolutionary Perspective: Inferences from Decorative Diversity and Interassemblage Distance in Illinois Woodland Ceramic Assemblages. *American Antiquity* 60:7–36.
- Netting, Robert M.  
1993 *Smallholders, Householders: Farm Families and the Ecology of Intensive, Sustainable Agriculture*. Stanford: Stanford University Press.
- Oren, S., and R. Schwartz  
1988 Diffusion of New Products in Risk-Sensitive Markets. *Journal of Forecasting* 7:273–287.
- Pinker, Steven  
1997 *How the Mind Works*. New York: Norton.
- Rabin, Matthew  
1998 Psychology and Economics. *Journal of Economic Literature* 36:11–46.
- Richerson, Peter J., and Robert Boyd  
1992 Cultural Inheritance and Evolutionary Ecology. In *Evolutionary Ecology and Human Behavior*. Eric A. Smith and Bruce Winterhalder, eds. Pp. 61–92. New York: Aldine de Gruyter.
- Rogers, Everett M.  
1962 *Diffusion of Innovations*. New York: Free Press.  
1995 *Diffusion of Innovations*. 4th edition. New York: Free Press.
- Rogers, Everett M., and D. L. Kincaid  
1981 *Communication Networks: Toward a New Paradigm for Research*. New York: Free Press.
- Ryan, B., and N. C. Gross  
1943 The Diffusion of Hybrid Seed Corn in Two Iowa Communities. *Rural Sociology* 8:15–24.
- Stigler, G. J., and G. S. Becker  
1977 De Gustibus Non Est Disputandum. *American Economic Review* 67(1):76–90.
- Tomasello, Michael  
2000 *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Tooby, John, and Leda Cosmides  
1992 The Psychological Foundations of Culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. J. H. Barkow, J. Tooby, and Leda Cosmides, eds. Pp. 19–136. New York: Oxford University Press.
- Waltman, P.  
1974 Deterministic Threshold Models in the Theory of Epidemics. Berlin: Springer-Verlag.
- Weibull, Jorgen W.  
1995 *Evolutionary Game Theory*. Cambridge: MIT Press.
- Whyte, William H.  
1954 The Web of Word of Mouth. *Fortune* 50:140–143.
- Wilbert, Johannes  
1993 *Mystic Endowment: Religious Ethnography of the Warao Indians*. Cambridge, MA: The President and Fellows of Harvard College.
- Young, H. Peyton  
1998 *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton: Princeton University Press.

## Appendix A

### Derivation of $P_1$ , $L$ , and $P_2$ from Figure 2

$P_1$ ,  $L$ , and  $P_2$  (from Figure 2) can all be related through the cumulative normal distribution using  $\mu$ ,  $\sigma^2$ , and  $d$ . If  $F(\mu, \sigma^2, x)$  represents the cumulative normal distribution evaluated at  $x$ , then

$$\begin{aligned} P_2 &= F(\mu, \sigma^2, -d - \mu) \\ P_1 &= 1 - F(\mu, \sigma^2, d - \mu) \\ L &= 1 - P_1 - P_2 = F(\mu, \sigma^2, d - \mu) - F(\mu, \sigma^2, -d - \mu) \end{aligned}$$

### Derivation of Equation 3

Equation 3 is a robust result of a variety of approaches to formalizing biased cultural transmission and replicator dynamics. Here I outline only one possible derivation. More extensive treatments can be found in Boyd and Richerson 1985, Cavalli-Sforza and Feldman 1981, Gintis 2000, and especially Weibull 1995.

**Table 1.** Probabilities of switching traits.

Naive's Current Trait	Model's Trait	Probability of Trait 1	Probability of Trait 2
1	1	1	0
1	2	$\frac{1}{2}\{1 + (r_1 - r_2)\}$	$\frac{1}{2}\{1 - (r_1 - r_2)\}$
2	1	$\frac{1}{2}\{1 + (r_1 - r_2)\}$	$\frac{1}{2}\{1 - (r_1 - r_2)\}$
2	2	0	1

In this two-trait formulation, the symbol  $q$  tracks the frequency of individuals with Trait 1, while  $1 - q$  tracks the frequency of individuals with Trait 2. Naive individuals enter the world and initially acquire the trait of their parents. Later in life, as adolescents, they pick an individual at random from the population and compare the  $r$ -value of this model (in the case of prestige-biased transmission) or the model's trait (in the case of direct bias) with their own  $r$ -value or the  $r$ -value of the trait they acquired from their parents. The probabilities of switching traits or keeping the current trait are shown in Table 1.

Remember,  $r$ -values—the replicatory propensities for each of the traits—may also contain a conformist component, which depends on the frequency of the trait in the current population. In this circumstance, when a naive individual encounters someone with a trait different from his or her own, he or she assesses the relative frequency of the two traits and uses this in his or her imitation decision. The  $r$ -values are described as follows:

$$r_1 = b_1(1 - \alpha) + \alpha(q - \frac{1}{2})$$

$$r_2 = b_2(1 - \alpha) + \alpha(1 - q - \frac{1}{2})$$

$$r_1 - r_2 = (b_1 - b_2)(1 - \alpha) + \alpha(2q - 1) = b(1 - \alpha) + \alpha(2q - 1)$$

As I explain in the main text,  $\alpha$  represents an individual's degree of reliance on conformist transmission versus his or her reliance on other forms of cultural transmission—it measures the strength of conformist transmission in the cognition of social learning. The variables  $b_1$  and  $b_2$  are meant to potentially capture a variety of other forms of non-frequency-dependent transmission biases, including direct and prestige biases. Direct biases capture the “catchiness” of an idea, belief, or practice. For example, if Trait 1 is a “belief in the afterlife” and Trait 2 is a “belief that death is the end,” then, for a variety of psychological reasons, we might suspect that  $b_1 > b_2$ . This means that either human brains in general or human brains acculturated in certain places are probabilistically more likely to adopt Trait 1 than Trait 2. Similarly, if Trait 1 is the behavior of “drinking sweet carbonated beverages” and Trait 2 is “drinking well water,” then, for entirely different psychological reasons, we might expect  $b_1 > b_2$  because the be-

**Table 2.** Frequency of possible pairings.

Possible Pairings	Frequency of Pairings
Trait 1–Trait 1	$q^2$
Trait 1–Trait 2	$q(1 - q)$
Trait 2–Trait 1	$(1 - q)q$
Trait 2–Trait 2	$(1 - q)(1 - q)$

havior of drinking sweet carbonated beverages is (for whatever reason) more catchy than the alternative. Or, under prestige-biased transmission, the  $b$ 's might represent an observable quality of the model, such as his or her prestige, skill, or age. We have reason to believe that  $b_1 > b_2$  whenever the model is considered particularly skilled, even when the skill has nothing to do with Trait 1 or 2 (Henrich and Gil-White 2001). In either case,  $b (= b_1 - b_2)$  provides the relevant comparison that determines both the direction and the rate of cultural evolution.

For conformist transmission, I have assumed that individuals can accurately determine the frequency of the more common trait. This is a fairly harmless assumption that substantially simplifies the mathematics. It turns out that whether individuals use a sample of only three models (picking three people at random from the population) or the entire population does not significantly change the qualitative results (Boyd and Richerson 1985:213). Further, this sample can be somewhat nonrandom without substantially changing the qualitative results (Boyd and Richerson 1985:211).

Using the frequency of each possible pairing (Table 2), we can calculate the frequency of Trait 1 after this imitation process by multiplying the frequency (or probability) of each pairing by the probability of ending up with Trait 1. We get the following recursion:

$$q' = q^2(1) + q(1 - q)\frac{1}{2}\{1 + (r_1 - r_2)\} + (1 - q)q\frac{1}{2}\{1 + (r_1 - r_2)\} + (1 - q)(1 - q)(0)$$

Simplifying this, we get an expanded form of Equation 3, showing the conformist component:

$$q' = q + (1 - q)q(r_1 - r_2) = q + q(1 - q)B$$

$$= q + q(1 - q)\underbrace{\{b(1 - \alpha)\}}_{\text{Other Biases}} + \underbrace{\{\alpha(2q - 1)\}}_{\text{Conformist Component}}$$

This corresponds to Equation 11 in the text. When  $\alpha = 0$ , which effectively eliminates conformist transmission and frequency-dependent biases, we arrive at Equation 3 in the main text.