The Gruter Institute Working Papers on Law, Economics, and Evolutionary Biology

Volume 2, *Issue* 1 2001 *Article* 3

Achievement Bias in the Evolution of Preferences

Edward Castronova*

*Cal State Fullerton, castro@indiana.edu

Achievement Bias in the Evolution of Preferences

Edward Castronova

Abstract

The paper develops an evolutionary selection model of the cultural transmission of preferences, focusing on the survival probability of certain preference types. The fitness of a preference is defined in terms of the ease with which its carrier can transmit the preference to the young. For example, a taste for work gives its carriers more income than is obtained by those who carry a taste for leisure. If higher income allows a given carrier to transmit her preferences more easily, then those with a taste for work will be more likely to transmit their preferences to the young; hence a taste for work will be more evolutionarily fit than a taste for leisure. In general, cultural transmission of preferences will favor any tastes that facilitate their own transmission, especially tastes for social achievements such as income, power, mass communication, and knowledge. The resulting pattern of tastes can be biased in the following sense: if the young generation were not influenced by achievement effects, they would choose preferences that would make them happier.

KEYWORDS: preferences, evolution, well-being

January 2002

Achievement Bias in the Evolution of Preferences

Edward Castronova
Associate Professor of Economics
Cal State Fullerton

ABSTRACT:
The paper develops an evolutionary selection model of the cultural transmission of preferences, focusing on the survival probability of certain preference types. The fitness of a preference is defined in terms of the ease with which its carrier can transmit the preference to the young. For example, a taste for work gives its carriers more income than is obtained by those who carry a taste for leisure. If higher income allows a given carrier to transmit her preferences more easily, then those with a taste for work will be more likely to transmit their preferences to the young; hence a taste for work will be more evolutionarily fit than a taste for leisure. In general, cultural transmission of preferences will favor any tastes that facilitate their own transmission, especially tastes for social achievements such as income, power, mass communication, and knowledge. The resulting pattern of tastes can be biased in the following sense: if the young generation were not influenced by achievement effects, they would choose preferences that would make them happier.

Acknowledgments: I am indebted to Nina M. Castronova and other members of the PhD Program in Marriage and Family Therapy at Syracuse University, whose ideas contributed significantly to the genesis and development of this paper. Thanks also to Timur Kuran, Paul Zak, Herb Gintis, and participants in seminars at USC, Claremont Graduate University, and Cal State Fullerton, for their helpful comments.

2.0

I. Introduction

This paper focuses on preferences as the object of evolutionary selection. It assumes that the fitness of a preference is determined by its ability to be transmitted from the members of one generation (usually a parent) to the members of the following generation (usually that parent's child). A specific 'preference' is a member of a set of utility function parameters, and it is assumed that young people obtain an initial parameter set, at the age of maturity, by a process that combines elements of biological hardwiring with elements of enculturation.[1] The focus here will be on the latter process, by which an individual's cultural environment affects his or her tastes. Increasingly, social scientists are unwilling to take preferences as given, and new lines of research are beginning to explore the role of the social, economic, and institutional environment in shaping tastes (Frank, 1987; Kuran, 1991; Boyd and Richerson, 1994; Bowles, 1998; Ben-Ner and Putterman, 1998; see also Becker, 1996; Gintis, 1972).[2] The endowment of preferences being, at least partly, a cultural process, it can be studied using theories of cultural evolution (Boyd and Richerson, 1985). Recent papers in the economics literature that study preferences strictly from the standpoint of cultural evolution include Rogers' (1994) analysis of time preferences, Bisin and Verdier's (2000) analysis of ethnic and religious traits, and Bowles and Gintis' (1998) study of the evolution of pro-social norms, and Guth and Ockenfels' (2000) work on cooperation.[3]

The contribution of this paper is to focus on the mechanisms by which preferences may be culturally transmitted, and how different kinds of preferences will endow their carriers (that is, the people who have these preferences) with different levels of resources that are effective in that mechanism. Preferences that are amplified by the

mechanism will be more fit and will flourish. The patterns of preferences that result are not necessarily the ones that maximize biological fitness (a point that has been made before), nor are they ones that maximize human happiness or well-being (a point that has not been made before). They will, however, share a general feature that I will call 'achievement.' I will argue that tastes for certain social achievements (such as income or knowledge) are systematically favored by the preference transmission mechanisms of human societies in general. As a result, there is a general bias in human culture in favor of tastes for these achievements, regardless of whether these activities increase our biological fitness or improve our psychological sense of well-being.

For example, consider a small hunter-gatherer village and imagine that all orphans are given to the leader to be raised. Suppose children start adult life with the preferences of the person who raised them. The transmission mechanism could be described as one in which children are usually endowed with their parent's preferences, but sometimes they may receive the preferences of the leader. Given this mechanism, a preference for power has an evolutionary advantage over preferences for other activities. People who enjoy the pursuit of power are more likely to become leaders; being leaders, they are able to pass their taste for power activities to the village orphans. If the leader represents x percent of the village population, we would expect at least x percent of the village's children to enjoy power-seeking activity simply because they are the leader's children (assuming leaders and others are equally fertile). However, if an additional y percent of the children are orphans who are raised by the leader, then the percentage of children who have a taste for power is x + y. Thus, since x + y is greater than x, the transmission mechanism favors power-seeking preferences. Moreover, in each generation, the fraction of children who

are power-seekers is higher than in the previous generation. Eventually, everyone in the village will, at the age of maturity, have a taste for activities that lead to power; everyone will want to be the leader, at least when they emerge from childhood.

Note that this example does not require that a taste for power should increase biological fitness (indeed, leaders and followers are assumed to have equal biological fitness). Nor is it necessary that the pursuit of power make people happy or satisfied.[4] If we assume that the pursuit of power in itself does not make people happy, then it is probable that many or most young adults, as they mature, would eventually remove the desire to be leader from their brains; they would engage in the time-consuming process of altering their own preferences.[5] Nonetheless, it is still true that the person who does become leader, happy or not, will generally have a stronger taste for power than anyone else, and will pass that taste on to the children she raises. In the steady state, this society would consist of repeated cohorts of universally avid power-seeking children, who then, as adults, become either successful but unhappy leaders, or unsuccessful but happy followers.

Of course, forces well beyond those present in this simple example will have an affect on cultural transmission, so the main objective of the paper is to show the existence of achievement bias in a more plausible general model of preference transmission. The model will take account of several realistic forces. First, it will account for the fact that culture is not everything, and that some element of human motivation is hard-wired in the brain. Second, it will allow people with the same tastes to group themselves, so that a child is less likely to be encultured by an adult from another group. Third, the model allows children some judgment, in that they will be more likely to adopt the preferences

of adults who are happier. Fourth, it will account for natural selection (the differential mortality rates of people with different characteristics) as well as endogenous preference changes during adult life. Fifth, it will allow group selection, as when groups with low resources are more likely to die out. The model does not explicitly account for assortive mating, since its force is already captured in the assumption that people with similar tastes tend to stay in separate groups.

As for methods, the paper adopts the position from the start that utility and well-being are not the same. Utility is a numerical ranking of desired states that serves as a guide to the rational choice of actions, while well-being is a substantive assessment of the goodness of a person's life. In one way of thinking about it, well-being is synonymous with happiness, and throughout the paper the two terms will be interchanged freely.[6] Clearly, happiness and utility are not always the same, since it is possible that rational actions that achieve higher-ranked and hence more desired outcomes may still leave the agent less happy. To view happiness and utility as distinct is not uncommon (and will be defended below), but it does introduce some methodological wrinkles that should be kept in mind. First, an individual who maximizes utility over choice of some good x will have a value x^* that represents the utility-maximizing choice, but also a different value x_0 that represents the choice of highest happiness. In a typical economics paper, one would discuss x^* as the value that is "best" in some sense. Here, however, it may or may not be best; x_0 is best, always. Second, most papers on the cultural evolution of preferences assume that agents will switch their type if they encounter other agents with higher payoffs, i.e. higher utility. Here, however, payoffs and happiness are not the same, and it will be argued that the preference changes should be based not on the comparison of

payoffs, but on the comparison of happiness levels. The reason is that happiness can be compared across people (with error of course), but utility cannot be.

The paper is organized as follows. Sections II and III make an argument for approaching cultural transmission as a matter of well-being, not utility. This is done in two parts, responding to two separate arguments in favor of deriving all cultural dynamics from utility. The first argument is that utility, as economists understand it, is essentially the same thing as biological or 'Darwinian' fitness.[7] Section II makes a case against that view. The second argument is that utility is essentially the same thing as happiness; Section III makes a case against that view and derives some implications for the way that one should model preference evolution. Section IV then identifies several achievement mechanisms and describes how they affect preference evolution. Section V illustrates how these mechanisms may function in a formal model of preference evolution, using tastes for work and leisure as an example. Section VI concludes.

II. Cultural Transmission: Utility and Fitness

The idea that culture can be understood through the lens of evolutionary selection was introduced in several seminal works (Ruyle, 1973; Boyd and Richerson, 1985; Cavalli-Sforza and Feldmann, 1981; Lumsden and Williamson, 1981). The basic insight is to treat a cultural characteristic (for example, that red octogonal signs mean 'stop') as a gene-like entity, variously called "meme" (Dawkins, 1976) or "culturgen" (Lumsden and Williamson) or simply "trait." The trait lives in an environment that consists of the brains of humans and the modes of interaction between them, and it may thrive or die there. A trait thrives if humans successfully teach it to other humans, especially children.

A core issue in the literature has been the question of whether cultural evolution must be consistent with biological evolution. Culture itself must be the product of natural selection, it is argued, and therefore culture must serve the biological fitness of human beings. If so, then culture is merely an intermediate object, a medium by which the forces of natural selection shape human society to serve the goals of biological fitness. Rogers (1994), for example, assumes that time preferences must be in an evolutionary equilibrium, in the sense that no change would improve Darwinian fitness. This then implies that the MRS in utility must be equal to the MRS in Darwinian fitness; in effect, if culture exists at all, it exists only to give people the time preferences that maximize their Darwinian fitness.

Boyd and Richerson (1985), however, show that while culture itself may be adaptive (i.e. may enhance Darwinian fitness), not all of the behavior it produces need do so. In general, a child will have a better chance of surviving to child-bearing age if it is capable of culture, but that capability, in certain environments, may generate cultures that encourage behaviors that lower Darwinian fitness.

There are two specific cases, however, in which cultural transmission perfectly preserves traits that serve Darwinian fitness. The first is *guided variation with unbiased transmission*: people learn various traits as they go along, and then transmit them to their children. I will refer to this as the 'learning' mechanism. The second is *transmission with direct bias*: people choose which traits to adopt according to some criterion, such as happiness. I will refer to this as the 'direct' mechanism.

If these are the only cultural transmission mechanisms, then cultural transmission perfectly preserves the pattern of traits that best serves Darwinian fitness. Take the direct

mechanism as an example. Suppose people choose traits using some criterion ('I choose to adopt my teacher's traits because she seems to be successful'). One may ask where the criterion (success is good) come from? Under natural selection, people would be more likely to survive to parenthood if they use choice criteria that serve fitness. Thus, evolutionary pressure molds our brains so that our notion of success is consistent with fitness, and we therefore, in choosing traits that make us successful, only choose traits that serve fitness. Hence, the direct mechanism is molded by natural selection. If natural selection would have dictated a distribution F(x) of some trait x in a population, then the distribution under cultural transmission with the direct mechanism will also be F(x).

However, other mechanisms of cultural transmission may encourage traits that do not enhance fitness, and indeed such maladaptive traits may survive. That this can happen is the result of identifiable biases in the way that culture transmits traits. Boyd and Richerson describe two such biased mechanisms: frequency dependent bias and indirect bias. Frequency dependent bias occurs because humans may be more inclined to adopt a trait that is shared by many other humans in their environment. Indirect bias stems from a tendency to imitate unimportant traits we can observe (such as a certain language dialect), on the assumption that they are correlated with important traits we cannot observe (such as business acumen).

That such cultural biases can move society well away from patterns of maximal biological fitness is illustrated by the demographic transition that accompanies industrialization (Boyd and Richerson, 1985, p. 200). In developed economies, people have fewer children, and that seems to be a biologically maladaptive behavior. However, it can be explained as the result of the indirect bias mechanism, combined with natural

selection. Small families are wealthier families, and members of wealthier families are a) more likely to be imitated by others, under the indirect bias mechanism, and b) more likely to survive to an age at which they can transmit culture.

For economists, this anthropological discussion brings up an important question: are preferences transmitted by fitness-preserving mechanisms? It certainly is possible; indeed Rogers (1994) explicitly assumes so. Suppose, for example, that people choose their preferences solely according to their payoffs in the environment, and then transmit them in a largely unconscious process to their children. That would be an example of the learning mechanism. Moreover, suppose that people tend to adopt the preferences of other people who seem to have higher payoffs (producing what is known as a 'replicator dynamic'). That would be an example of the direct mechanism.[8] From the discussion above, we know that these mechanisms would ensure that the preferences of agents would exactly reflect Darwinian fitness; utility maximization would be the same thing as fitness maximization. Gintis (2000) reviews a very large literature (mostly in economics) in which preferences evolve in this way, typically through the replicator dynamic. Thus, in this literature, the implicit assumption is that utility and fitness are equivalent.

The broader implications of the anthropological literature, however, are that this assumed equivalency between utility and fitness is not always warranted. Suppose that in some situation the dominant form of cultural transmission of preferences involves frequency dependent bias (e.g. herd-like behavior, informational cascades, or reputational cascades) instead of direct bias (e.g. the direct mechanism and the replicator dynamics). Then we know from the anthropological literature that the resulting pattern of preferences need not give the population the highest level of Darwinian fitness (Gintis 2000, pp. 217-

219). [9] Moreover, in models with these herd-like evolutionary mechanisms, it can be the case that types with low payoffs can survive (Bikhchandani, Hirshleifer, and Welsh, 1992; Harrington, 1999; Banerjee 1992). Furthermore, since now the survival of preferences is not dictated entirely by natural selection, it is no longer necessarily true that each action that raises utility also raises fitness. The two objectives, fitness and utility, are no longer equivalent.

This does not mean, of course, that biological fitness has nothing to do with preferences at all. There is considerable evidence that emotional states are to some extent hard-wired into the brain (Miller, 2000; Barkow, Cosmides, and Tooby, 1992; Ledoux, 1996). At the same time, it is a core assumption of economics that individuals pursue happy states and avoid unhappy ones. Taken together, these two ideas suggest that tastes are partly determined by our desire for happiness, and that the states we consider 'happy' are those selected by millennia of biological evolution. Thus, biology has some influence on our preferences. That influence may not be all that there is however, if we believe the cultural anthropology literature. There may be cultural forces that influence tastes, in a way that has nothing to do with the pursuit of happiness.

The case that all our cultural traits must be consistent with biological fitness seems unpersuasive on a deeper level as well. Four million years have elapsed since the appearance of the first hominids, and the amount of time in which these organisms have been capable of culture (dating from the first stone figurines) is only about 32,000 years. In other words, we have been watching an opera for two hours, and cultural humanity has just come on stage and sung for about one minute. We cannot tell from her performance in that brief time whether she will still be alive when the curtain falls. True, cultural

humanity seems to play a dominant role at the moment. But given the short amount of time in which we have been capable of culture, it seems specious (to me at least) to argue that this or that cultural practice must be or must not be adaptive in the biological long run for the current version of hominids.

Thus, for several reasons, the anthropological literature suggests that it is not appropriate to assume that preferences in economic models must also serve Darwinian fitness. Instead one must begin by addressing the question of how well the preference can survive in its evolutionary environment, i.e., in the minds of the people who carry and transmit it. This requires an examination of the mechanism of transmission. If a preference is transmitted by things like the learning mechanism or the direct mechanism only (as with a replicator dynamic operating on payoffs), then the frequency of the preference in the population will be the same as if that frequency were instead dictated only by Darwinian fitness. If the preference is transmitted by some other mechanism, however, its frequency will be not be dictated by Darwinian fitness.

III. Cultural Transmission: Utility and Well-Being

The preceding discussion opens the possibility that a force unrelated to Darwinian fitness, and unrelated to emotional states of happiness or well-being, may drive the evolution of preferences.[10] To pose this question is to permit utility and well-being to be different, a distinction that is not common for economists to make. It is worth reflecting on the case for treating the two notions as less than completely equivalent, which is fairly well-developed in the philosophy of economics. The distinction has important implications for the modeling of preference change.

To begin with, utility is just an enumeration of the extent to which the agent has achieved his goals, while well-being is a substantive notion about a person's state of mind or existence (Hausman and McPherson, 1996, ch. 6). Specifically, there is nothing in the theory of utility requiring that preference satisfaction necessarily lead to emotional satisfaction. Some of the things you desire are not good for you; the attainment of your goals will not always make you happy. Indeed, studies of subjective happiness (Frey and Stutzer, 2000; Kahneman, Diener, and Schwarz, 1999) suggest that a significant increase in income ceteris paribus, which will raise utility in almost every reasonable economic model, is usually not matched by a significant increase in happiness.[11] Several studies have shown that long-run increases in per capita GDP do not significantly increase satisfaction (Easterlin, 1974, 1995; Blanchflower and Oswald, 2000).

The distinction between utility and happiness has fairly important implications for models of cultural transmission of preferences. In most of the endogenous preference literature (see Gintis, 2000), the transmission mechanism is of the direct type, relying on comparisons of payoffs between agents. Agents of type A who perceive that agents of type B are 'doing well,' i.e. obtaining high payoffs, are likely to change their type from A to B. But why would an A type desire the payoffs of a B type? Perhaps it is because the payoffs represent substantive well-being. In other words, suppose we assume that the A types believe the following: if they were to adopt the preferences of the B types, and then act as those preferences dictate, they would find themselves with higher payoffs and also higher well-being.

Such an assumption may make sense in certain circumstances but is not tenable if utility and well-being are allowed to be different. Suppose the B types happen to care

little about their freedom while the A types care quite a bit about it; otherwise they are the same. Suppose further that in this society, slaves make a great deal of money but free people make very little, so that the B's generally have higher utility - they make a lot of money and do not really care about being slaves. Suppose still further that an A type always has lower utility when a slave than when free. And finally, suppose that slaves always score lower than free people on psychological tests of subjective happiness. According to the payoff based cultural transmission mechanism, A types will say 'were I to adopt the B's apathy regarding freedom, I could let myself become a slave, make more money, and obtain the higher utilities of the B types.' If an A type did this, however, and became a slave, in what sense is he better off? True, his utility, being now determined by the B preferences, is higher. However: a) his utility under the A preferences is now lower, and b) he is less happy, according to psychological tests of emotional satisfaction. Thus, cultural transmission according to payoffs involves assuming that people would willingly choose to make a change that, according to their preferences at the time of the choice, would lower their utility. And of course it also assumes that people would make a change that would lower their level of emotional satisfaction or happiness.

The example highlights two problems with the assumption that people will switch preferences according to the utilities that those preferences allow. The first problem derives simply from the fact that utility cannot be compared across people. Utility is just an index number, an ordering of states. To say that Bird has reached a state numbered 107 for him while Castronova has reached a state numbered 2.38769384 for him is not to say that Bird is in a better situation. It certainly does not follow that Castronova would want to *be* Bird, as payoff-based cultural transmission requires.

The second problem derives from the fact that satisfying preferences does not guarantee higher levels of human well-being. People willingly do things that make them unhappy, or that are not good for them in some broad sense. Otherwise self-command would not be the important research topic that it is in many fields -- including economics to an increasing degree (Thaler and Sheffrin, 1981; Schelling, 1984; Becker and Murphy, 1988; see the chapters in Kahneman, Diener, and Schwarz, 1999, Part V).[12] The concept of utility is a proper guide to the analysis of actions, but it is no more than that; it is not a perfect guide to the substantive value of actions; it does not indicate how happy one will be from taking those actions; it does not measure how much a person's world is improved by those actions. Sen (1993 and elsewhere) and others (see Hausman and McPherson, 1996) have persuasively argued for substantive concepts of well-being, ones that are not based on preference-satisfaction alone. Again, to say that Bird has higher utility than Castronova is not, in general, to say that Bird has a better life than Castronova, or is happier. And again, it certainly does not follow that Castronova would want to *be* Bird, as payoff-based cultural transmission requires.

The possibility that preferences may evolve according to something other than utility has been recognized in the literature, although the implications for the substantive theory of well-being have not been explored to any great degree. Guth and Yaari (1992), Guth and Ockenfels (2000) and others have developed models of indirect preference evolution, in which the fitness of a type is determined by a subset of total payoffs. Thus, if utility is given by $u = x + y$, fitness is given by $f = x$. Types replicate more rapidly if their actions, which are motivated by both x and y, happen to lead to higher values of x. Because x is an element of utility, however, such models are essentially utility based,

payoff-based evolution models. Moreover, they make no distinction between utility and well-being. Indirect preference evolution models are therefore not immune to the discussion above.

One might respond by arguing that there may be some metric, such as money, that is comparable across people and is reasonably related to well-being. If payoffs are defined as money, however, a payoff-based mechanism of cultural transmission assumes that type A agents would switch to type B if and only if type B agents had more money. Thus it assumes that the types care only about money and nothing else. In other words, it assumes that money is not just a reasonable metricof well -being: it is the only metric of well-being. In some circumstances such an assumption may be reasonable, but in general it is not. In fact, assuming that cash is equivalent to well-being is more restrictive and unreasonable than the prior assumption that utility is equivalent to well-being. Denoting payoffs as cash may solve the comparability problem, but only by making more heroic assumptions about the nature of well-being.

Evidently, models of the cultural transmission of preferences can be based on payoff comparisons only under certain circumstances. What are those circumstances? When is it plausible that a type A would *want* to be a type B? The most immediate and intuitive answers generally involve well-being. That is, we would expect type A to desire type B's situation if and only if A believes that B has higher well-being, in some substantive sense. Perhaps B appears to be happier, or B scores higher on psychological tests of emotional satisfaction. Perhaps the rate of suicide and depression is lower among type Bs. Perhaps the Bs all live the Good Life according to some objectively substantive criterion (e.g. Sen's functionings): their health is better, their family relations are more

peaceful, they consume more art, etc. All of these circumstances are probably hard-wired by our mental biology; they are comparable (albeit with error) across people; and it is more than plausible that people with low well-being (by these definitions) would want to be people with higher well-being. Intuition alone almost forces one to conclude that well-being, and not utility, must be the fundamental criterion by which people choose their preferences.

If well-being is the fundamental criterion of preference evolution, then, when will it be justifiable to assume that utility and well-being are equivalent, so that cultural transmission can be based on payoffs? Such an assumption is justified in certain circumstances if one now reconsiders the role of Darwinian fitness. Suppose first that well-being is essentially the mental state of happiness.[13] Suppose that biological evolution endowed the human mind with the emotions of happiness and sadness simply as a way to motivate survival.[14] That is, all situations produce good and bad emotions depending entirely on whether Darwinian fitness is rising or falling. Since people are driven to seek happiness and avoid sadness, they are driven by this psychological pattern to seek fitness and avoid extinction. In the course of their lives people will form certain preferences that effectively dictate their behavior; they learn to do this and not that because doing this makes them happy, or because someone else did this and became happy. In other words, they apply a criterion of substantive well-being, in order to learn which preferences are best (the learning mechanism), as well as to copy preferences from others (the direct mechanism). From the anthropology literature discussed in Section II, we know that these mechanisms will simply preserve the preferences that best serve the criterion by which the preferences are chosen. That is, if people choose their preferences

according to criterion X, these mechanisms will ensure that the population will consist entirely of people whose utilities are maximized when X is maximized. Here the criterion X is happiness. Hence, if learning and direct mechanisms are the sole source of cultural transmission, then utility will be the same thing as happiness. Moreover, biological evolution is assumed to have given people the criterion X so that it accords with Darwinian fitness; natural selection has made happiness and fitness equivalent. Thus biology makes happiness and fitness equivalent, while culture makes happiness and utility equivalent; hence fitness, happiness, and utility are all the same.

The general equivalence of these three depends entirely on the structure of cultural transmission, however. It holds only because natural selection translates fitness into happiness, and then culture translates happiness into utility. However, in Section II it was argued that cultural transmission does not always have this effect. That is, suppose that happiness is indeed the criterion by which people choose preferences under the learning and direct mechanisms. If these are the only operating mechanisms of cultural transmission, then preferences will be maximized when happiness is maximized. However, there may be other mechanisms at work, such as herd behavior and the like. These generate evolutionary forces that push preferences away from the criterion that is employed in the learning and direct mechanisms. Naïve agents who are searching about for a preference to adopt will be subject to conflicting forces: on the one hand, they are more likely to adopt (or be encultured into) the tastes of someone who seems happy; on the other, they are more likely to adopt (or be encultured into) the tastes of people whom they see more often – whether or not those people are happy. As a result, the preferences that are adopted may not maximize happiness. When the learning and direct mechanisms

are in conflict with other mechanisms, the criteria they employ will not be the only determinant of preferences. Thus, utility and happiness will not be the same, and it follows that utility and fitness will not be the same.

The end result of the argument in these two sections is this: cultural transmission of preferences does not guarantee that preferences will serve either Darwinian fitness or human well-being. Everything depends on the nature of the mechanisms by which preferences are culturally transmitted. If preferences pass from mind to mind through a mechanism that relies *solely* on a criterion of well-being, then culture will translate the well-being criterion into a utility criterion. If preferences pass from mind to mind through some other mechanism, then the well-being criterion may not be translated directly into a utility criterion. Utility may be maximized when well-being is not. The distinction between utility and well-being may be more than a philosophical nicety, it may be a core element of human cultural existence.

The question now becomes, what mechanisms transmit the preferences that are of greatest interest to economists? Virtually all the attention in the economics literature has focused implicitly on the learning and direct mechanisms, probably because these have the happy property of producing equivalence between utility, well-being, and fitness. What other mechanisms are worth examination? Can one make the case that the most important preferences, such as those for work, leisure, income, offices, and so on, are more likely to be transmitted by the direct and learning mechanisms alone? If so, then we can conclude that the pattern of development of these tastes over time generally coincides with an increase in the substantive well-being of the human species. If not, then we face

the possibility that the development of preferences has not necessarily coincided with an increase in the substantive well-being of the human species.

IV. Achievement Mechanisms

Utility, happiness, and fitness all coincide when culture transmits preferences solely according to happiness. Some cultural mechanisms, however, may move preferences from one person to another independently of either's happiness. In this section I argue that there is an important mechanism of cultural transmission that is distinct from the direct and learning mechanisms, and does not rely on comparisons of well-being to propagate in a population: Achievement. Achievements include things like social status, expertise, interaction, fame, and competitive success. This mechanism transmit preferences by a social achievement effect: all else equal, naïve agents (meaning those with as-yet unformed tastes) are more likely to adopt the preferences of successful people in society. Achievements in human society give those who have achieved a disproportionate influence on the processes of enculturation; this then creates a bias in the enculturation process towards tastes for achievement. Specific examples are described below.

Social Status. A taste is encouraged by status achievements if agents are more inclined to adopt the taste if its owner has high status in the social system. It is axiomatic among anthropologists, sociologists, and psychologists that people tend to imitate those who have social prestige. If such a status mechanism is operating in a culture, it means that naïve agents look to those with high status or prestige and tend to mimic their preferences. They may be inclined to adopt the preferences of happier people, all else

equal, but they also are inclined to adopt the preferences of more prestigious people, all else equal. It may be the case that they would choose the preferences of a comparatively unhappy yet prestigious person if the prestige level is sufficiently high.

As an example, suppose there are two brothers who are exactly alike except that one has an extremely intense love of public speaking, and the other has an extremely intense love of solitude and silence. The first will pursue opportunities to speak with all resources at his disposal; the second will avoid speaking to others at all costs. Given their efforts, it is reasonable to assume both are largely successful in their pursuits and manage to structure their lives according to their wishes: the public brother has an illustrious speaking career, the private brother becomes an utterly unknown and isolated man. Both are blissful. As a result, naïve agents will be inclined to copy the preferences of both brothers. The public brother, however, has far more social status and prestige. If status affects the adoption of tastes, the public brother will pass his tastes to more naïve agents than the private brother does. Thus, the taste for oratory has an evolutionary advantage over the taste for silence. The status mechanism gives it extra weight in transmission. In the next generation, their will be more people with a taste for oratory than if happiness were the only criterion by which naïve agents chose.

Expertise. Expertise achievements encourage a taste if agents are inclined to adopt the taste when its owner seems to have more understanding of important things. Under the expertise mechanism, naïve individuals tend to imitate those who seem to be knowledgeable. Suppose, for example, that a woman in a village has an intense desire to perform experiments with plants. In pursuing these interests, she obtains expertise in treating illnesses with plant-based compounds. If the expertise mechanism is operating in

this society, the fact that she is now perceived as knowledgeable about an important thing, medicine, will cause some naïve agents to copy her preferences (whether they find her happier or not). Thus, the expertise mechanism favors a taste for acquiring knowledge, at the expense of a taste for remaining uninformed.

Interaction. Interaction achievements encourage a taste if agents tend to adopt the tastes of those whom they encounter more often. Under this kind of mechanism, a naïve agent may put disproportionate weight on the fact that a certain type is encountered more frequently in her culture. As an example, suppose that a village of 100 people has 80 people with a strong taste for hunting and 20 people with a strong taste for farming. If children are encultured only by their parents or by adults chosen at random, 80 percent of each generation will be hunters. Suppose, however, that those with a taste for hunting go off and hunt for months, while those with a taste for farming stay home and farm. As a result, there are 20 farmers and only 10 hunters in the village at any one time. Children, who always stay home, are exposed to a non-random grouping of adults for enculturation, and the groupings systematically favor the taste for farming. As a result, more than 20 percent of the next generation will have a taste for farming. The interaction mechanism will encourage a preference for any activity that increases the rate of contact among people. It will discourage tastes for solitude.

Fame. Achievements of fame encourage a taste if carriers of that taste are able to project their personality more broadly, and if other agents are inclined to adopt the tastes of those who are known to many others. Fame differs from interaction in that interaction measures how often a person encounters others in two-way interactions, whereas fame measures how many people receive a person's one-way broadcasts. The fame mechanism

would favor a taste for public speaking in much the same way as the status mechanism does. If naïve agents tend to mimic the tastes of those who are famous (whether or not those people are happy or prestigious), the fame achievement will encourage the spread of tastes for activities that generate fame. It would not encourage silence.

Competitive Success. Winning a compe encourages a taste if agents are inclined to adopt the tastes of people who have emerged from a competitive process to obtain certain offices, and victory is regulated by the taste itself. Cavalli-Sforza and Feldman (1981) construct a theory of oblique cultural transmission with agents competing to enter offices through which they may enculture the children of others. Boyd and Richerson (1985, p. 179ff) give an example in terms of abstract reasoning skills. Suppose there is a town where teaching jobs are scarce and pay very well, so that there is intense competition to become a teacher. Moreover, suppose that candidates must pass a difficult test to obtain a teaching job, and that the test has a strong algebra component. Mr. Castronova happens to have a strong taste for algebra, and so he does well on the test and becomes a teacher. Being a teacher, Mr. Castronova now has disproportionate influence on the enculturation of the next generation. If students tend to imitate their teachers' preferences with greater frequency than they imitate the preferences of, say, meat inspectors, then the teachers' disproportionate love of algebra will translate into a disproportionate love of algebra among the children. Thus, competition encourages preferences that help people win competitions for those offices that heavily influence culture.[15]

Each of these examples emerges from the same basic logic: cultural evolution will favor preferences that can more easily propagate themselves in human minds. In the

jungle of human minds, the entity "a taste for oratory" is better able to survive than the entity "a taste for silence." The former is more fruitful than the latter; it presents itself with far greater frequency to the population of naïve and unencultured carrier minds. These more extensive contacts generate more extensive conversions among the carriers, and thus the taste for oratory reproduces itself more rapidly.

It is important to recall that the functioning of an achievement mechanism is distinct from considerations of human well-being. The demagogue may be a deeply unhappy man, but he is heard; no one hears the voice of the lonely hermit, happy though he may be. In a world without achievement mechanisms, the happiness of hermits would eventually make a taste for solitude dominant in the population. With achievement mechanisms, however, the hermits' taste for solitude will dominate only to the extent that their joy, in its intensity, can overcome the status and fame of the demagogues. In some cases it will do so, and all the population will love solitude. In other cases it will not, and the population will have a mix of tastes, or perhaps everyone will have a taste for demagoguery. Achievement mechanisms do not guarantee the dominance of the tastes they favor; they do make such dominance possible, however.

Are achievement mechanisms independent of Darwinian fitness? What explains their existence? According to one argument already given, we cannot really know whether any feature of culture is biologically adaptive. Common experience suggests, however, that status, fame, competition and the like are nearly universal attributes of human societies, from the hunter-gatherers on up to the digital proto-societies that thrive in cyberspace. Achievement mechanisms may or may not serve Darwinian fitness, but they do seem to exist in most human societies.[16] The point here is only that where they

exist, they will influence preferences in a way that is both inherently interesting for its effects on well-being, as well as distinct from the mechanisms that rely on payoffs that now dominate the literature.

V. The Evolution of Tastes for Work

This section presents a simple model of cultural evolution that illustrates the relationship between the achievement mechanism, utility, and well-being. Suppose we have a society of N individuals, indexed i = 1, … N, each having a happiness function

(1) $$\ln h_i = \alpha_0 \ln y_i + (1 - \alpha_0) \ln L_i$$

where h is happiness, y is income, L is leisure time, and α_0, which lies between zero and one, is a parameter showing how the mix of leisure and income translates into human happiness. We assume that α_0 is the result of biological evolution; every person in society is hardwired to be happiest when equation (1) is maximized.

Leisure and work must be chosen with respect to the following budget constraint, assumed to be the same for all people:

(2) $$y_i + wL_i = wT$$

where w is the wage rate and T is the time endowment. We will normalize the problem so that T = 1. Hence

(3) $$y_i + wL_i = w$$

Under these assumptions, the bundle $L^*_0 = 1-\alpha_0$ and $y^*_0 = \alpha_0 w$ will result in the maximum happiness for each agent.

Suppose, however, that each person's actions are determined by a different function, namely, the utility function:

(4) $$\ln u_i = \alpha_i \ln y_i + (1-\alpha_i)\ln L_i$$

where u is utility and α_I reflects individual i's taste for income relative to leisure. Each person chooses leisure and work to maximize (4) with respect to the budget constraint. Hence the utility-maximizing bundle is $L^*_i = 1-\alpha_i$ and $y^*_i = \alpha_i w$. This is the bundle that agents actually choose. It would be the happiness-maximizing bundle if and only if the agents happened to be endowed with the preference parameter $\alpha_i = \alpha_0$. The research question thus boils down to whether there are cultural processes that endow some or most agents with a preference parameter other than α_0. If not, then culture ensures that happiness and utility coincide. If so, culture ensures that happiness and utility do not coincide; our culture induces us to pursue goals that will only make us unhappy.

Suppose there are two types of preferences in the society, type g (grasshoppers) and type a (ants), with $\alpha_g < \alpha_a$. As in the fable, ants work harder and have higher incomes than grasshoppers.[17] Define the *misery index* m_i as follows:

(5) $$m_i = h(y_0, L_0) - h(y_i, L_i)$$

The misery index measures how unhappy individual i becomes when pursuing her goals, as defined by her utility function. For example, suppose $\alpha_0 = 0.5$, so that $y_0 = 0.5w$ and $L_0 = 0.5$. A person with $\alpha_i = 0.5$ would choose these bundles and would achieve maximum happiness. A person with $\alpha_i = 0.75$, however, would choose $y_i = 0.75w$ and $L_i = 0.25$ and would not achieve maximum happiness. We quantify the distance between the latter person's happiness and maximal happiness as m_i = [0.5ln(0.5w) + 0.5ln(0.5)] – [

$0.5\ln(0.75w) + 0.5\ln(0.25)$]. Persons with high values of m are assumed to be unhappy and have low levels of well being. Furthermore, the misery index is assumed to be comparable across people, so that if one person has a higher value of m than another, that person is assumed to be less happy. For concreteness, let $y_0 = \alpha_g w$ and $L_0 = 1 - \alpha_g$, so that the grasshoppers attain maximum happiness at the maximum of their utility. Hence, $m_g = 0$ by assumption, while $m_a > 0$. There is no loss of generality by this assumption; one could just as easily assume that ants were generally happier than grasshoppers and analyze cultural selection on ants.

A. Cultural transmission based on well-being alone

From the standpoint of well-being, it would be desirable if the cultural transmission mechanism were to favor grasshoppers over ants. Grasshoppers are happier. To specify the cultural transmission process, consider the following life cycle for each person. A person is born to a single parent and raised by that parent.[18] At the age of maturity, the young person is assumed to have the parent's preferences with probability 1 - p. With probability p, the young person is considered to be still naïve after parental influence, and will be encultured by someone other than her parent.

Assume the society is divided into two groups according to type, so that a person raised by a grasshopper, but not yet encultured as a grasshopper, is more likely to be exposed to grasshoppers than ants in the wider world. Let the fraction of the population who are grasshoppers be denoted s. Let the probability that someone raised in one group encounters an adult from another group be denoted δ, so that the probability of a naïve ant child meeting a grasshopper adult would be δs, with $0 < \delta < 1$. (Later we will relax

the assumption the both groups have the same values of p and δ). Similarly, the probability that a naïve grasshopper child meets an ant adult is $\delta(1\text{-}s)$.

Suppose that the enculturation process is as follows: a naïve child meets an adult of a given type with the probabilities above, and will adopt that adult's value of α with a probability that depends on the adult's characteristics. Otherwise the child adopts the other value of α. Again for concreteness, assume that children who encounter adults whose type is the same as their parents will adopt their parent's type with probability 1. Thus, a child will switch only if she encounters an adult of different type. Let the probability that a grasshopper child who meets an ant adult will adopt the α_a preference be given by the following formula:

(6) $$\pi_a = \Pr(ant \mid grasshopper) = \pi(m_a - m_g)$$

where $\pi' < 0$. The corresponding probability for conversion to grasshopper is $\pi(m_g - m_a)$. (In simulations will we assume that π is a linear function of the misery difference. As a result, with probabilities being bounded below at zero, we will be assuming that π is zero when the other type is less happy than the child's parent's type. Conversions into the relatively unhappier type never happen, and conversions into the relatively happier type are less likely if the misery difference is small.)

So far we have established a framework that is similar to the payoff-based evolution models that are common in the literature. Indeed, if equation (8) were defined in terms of payoffs instead of well-being, we would be able to derive a replicator dynamic: types with higher utility would grow in the population, while those with lower utility would die off (see Weibull, 1995). Here, however, the probability of switching

depends on well-being, and there is no inherent connection between the payoffs and the criterion of switching.

The dynamic that emerges from (8) can be derived as follows. Let s_0 denote the fraction of the population who are grasshoppers in generation 0. The fraction in generation 1, s_1, is determined by summing four factors:

- The percentage of grasshoppers who are not exposed to anyone but their parents is $(1-p)s_0$. This leaves ps_0 grasshopper children to be encultured by others.
- Of them, a fraction $1 - \delta(1 - s_0)$ encounter other grasshopper adults and receive the α_g preference.
- The remaining$\delta (1 - s_0)$ of grasshopper children who are encultured by other adults are encountered by adult ants. The probability that they switch from grasshopper to ant is given by π_a. Hence the probability that they will not switch, but will retain the grasshopper preference, will be $(1 - \pi_a)$.
- Finally, of the $1 - s_0$ ant children, a fraction 1 - p will be encultured as ants by their parents. The remaining $p(1 - s_0)$ ant children will encounter grasshopper adults with probability δs_0. The probability that they will switch from ant to grasshopper is given by π_g.

The frequency of α_g in generation 1 depends on its frequency in generation 0 as follows:

(7) $$s_1 = (1-p)s_0 + ps_0(1-\delta(1-s_0)) + ps_0\delta(1-s_0)(1-\pi_a) + p(1-s_0)\delta s_0\pi_g$$

If we define $\dot{s} = s_1 - s_0$ as the growth in s per generation, we can simplify (9) to

(8) $$\dot{s} = s_0(1-s_0)[p\delta\pi_g - p\delta\pi_a]$$

Define the term $B = p\delta\pi_g - p\delta\pi_a$ as the *net conversion rate*, the rate at which children are converting from ant to grasshopper ($B > 0$) or from grasshopper to ant ($B < 0$). The net conversion rate decreases as parents have more influence over their children's culture (p) and as groups are more insular (δ).

In the terminology of evolutionary theory, the system is at an evolutionary equilibrium at a point s^* if s^* is an asymptotically stable fixed point of (9) (Gintis, 2000, p. 173). Equation (10) reaches a fixed point when $\dot{s}$ is zero, at $s_0 = 0$ and $s_0 = 1$. When s_0 is not equal to 0 or 1, $\dot{s}$ is positive if and only if $B > 0$. In this case, $s_0 = 1$ is an asymptotically stable fixed point while $s_0 = 0$ is an asymptotically unstable fixed point; the only asymptotic equilibrium of the system is $s = 1$. If $B < 0$, then $\dot{s} < 0$ whenever s_0 lies between 0 and 1, which would make $s_0 = 0$ the asymptotic equilibrium.

Since we have assumed that children choose preferences according to happiness, and grasshoppers are happier, it will be the case that $\pi_g > \pi_a$ and therefore $B > 0$. The point $s^* = 1$ is asymptotically stable and is therefore the evolutionary equilibrium for the system. In other words, the fact that grasshoppers are happier than ants means that children systematically choose to adopt α_g over α_a. The frequency of grasshoppers grows with each generation. Over time, theα_a preference will eventually die out. Even if the rate of enculturation by non-parents is very small, and both communities are very insular (so that p and δ are small), the eventual extinction of the unhappy ants is assured.

Figure 1 illustrates the dynamics. It is based on a simulation of the model with a set of standard parameters (some of which have been set to zero and will not be discussed here, but are introduced in later sections). The graph shows values of $\dot{s}$ ("Change in s") as a function of the current level of s ("s0"). The simulation assumes that $p = \delta = 0.3$,

which implies that, if each group makes up 50 percent of the population, a child from one group will be exposed for enculturation to an adult of the other group with probability (0.3)*(0.3)*(0.5) = 0.045. That is, more than 95 percent of the children from each group will never face the possibility of switching type; which is to say, this simulation assumes only a minimal amount of cross-group enculturation.

The figures in the graph assume that α_g = ¼ and α_a = ¾, as well as w = 0.4. Grasshoppers prefer more leisure and less income than ants do. Because we have assumed that well-being is highest at the grasshoppers' mix of work and income, the misery index for grasshoppers is 0 by assumption, while for ants it is $m_a = 0.19$. The simulation further assumes that $\pi_i = -q_h*(m_i - m_{-i})$, where m_{-i} is the misery index of the other type and the parameter q_h is set to the value 3. If-$q_h*(m_i - m_{-i}) < 0$, π_i is assumed to be zero. Hence, we have $\pi_a = 0$ and $\pi_g = 0.57$. The net conversion rate thus favors grasshoppers, with B = 0.05172. The point at issue is whether the preferences of the unhappy ants can survive the process of cultural transmission.

In the figure, the $\dot{s}$ function has zeros at two points, s0 = 0 and s0 = 1. Between these two values, the function is always positive, meaning that s is growing. This implies that the system is unstable at s0 = 0: when everyone is an ant, a mutation that produces a small number of grasshoppers will not die out. The higher well-being of grasshoppers will cause the conversion of some number of ant children, and the process will continue until everyone is a grasshopper. Conversely, if everyone is a grasshopper (s0 = 1), a mutation that produces a small number of ants will die out as the unhappy ant children are systematically converted to being grasshoppers, with the system eventually returning to s0 = 1.

This example illustrates the possibility that cultural evolution can make utility and well-being equivalent. Because preferences are chosen only according to a well-being criterion, the preferences that eventually survive are those whose maximization also maximizes well-being. If happiness is the selection criterion, then cultural evolution ensures that utility – the guide to behavior– will also be the guide to happiness.

B. Cultural transmission with achievement bias

Suppose now, following the argument in Section IV, that cultural transmission is affected by achievement mechanisms in addition to the simple enculturation process just described. For concreteness, assume that the status mechanism has some influence on the way that preferences are transmitted. This would mean that young people, when deciding whether to be ants or grasshoppers, would pay attention not only to the happiness of their cultural models but also to their status. In the context of the model here, we will assume that income is the metric of status: people with more money have more prestige, so people with more money are more likely to enculture the naïve with their tastes.

Thus, in the model, a child adopts her parent's tastes with probability 1 – p and is exposed to other adults with probability p. Among these other adults, the unencultured child encounters someone from the other group depending on the population frequency of that group, as well as a group isolation factor $\delta < 1$. If the unencultured child from group -i does meet an adult from group i, the probability that she adopts the tastes of that person is π_i, and let it be a function of the incomes of the two types as well as their happiness:

(9) $$\pi_i = q_y(y_i - y_{-i}) - q_h(m_i - m_{-i})$$

where $q_y > 0$ and $q_h > 0$ are weights defining the relative influence of the income and happiness components, respectively, on the adoption probability.[19] The assumption that income influences the adoption of tastes ($q_y > 0$) can be justified empirically, since, in virtually all human societies, people with more money have more prestige, and prestige affects the adoption of tastes.[20]

There is a much broader justification, however. Even if some societies do not confer prestige on the rich, prestige exists in all societies and can be obtained through effort. This model describes how people make choices of effort toward some intrinsically rewarding goal, denoted y, and then asks what happens to the preferences of the young when those who achieve that goal, y, also receive prestige or any other resource that may make them more attractive as models for the young. It is not necessary, but rather seems to make the most intuitive sense, to think of the goal "y" as income.

One can further justify a focus on income because it is relevant for almost any achievement mechanism one can imagine. Income seems to raise status, as assumed above; it also seems to increase fame, the frequency of face-to-face interactions, one's ability to occupy competitive offices, and even expertise.

As for fame, in most societies the views and opinions of a person are more likely to be broadcast to groups of others, ceteris paribus, if that person is wealthier. The typical college professor has often sat through lengthy speeches of wealthy donors to the graduating class; how often has the speaker been poor? In post-industrial societies, the wealthy are the focus of intense media scrutiny, and for them, obscurity has become an extremely precious good. It is the CEO who speaks to the assembled employees, not the mail clerk, and the national evening news is written, edited, and read by people whose

incomes are well above average. Artists choose whether to produce for themselves or the masses; those who produce for the masses have more money (Cowen and Tabarrok, 2000). There is almost certainly a positive correlation between income and fame.

As for interactions, achieving high income typically makes a person commit herself to a life of social interactions: networking, career-building, and mobility. True, some people make a good deal of money by writing screenplays or playing chess on their home computers, but on average, a person has to socially active to be wealthy. Studies of social capital emphasize the social isolation of the poor (Wilson, 1987; Putnam, 1995). Career success involves a willingness to move geographically and also socially, so that the typical wealthy person will have moved through many more social circles than the typical poor person. The internet greatly facilitates personal interactions, and the typical internet user is wealthier than the typical non-user. Income and interaction go together.

As for competition mechanisms, gaining an office under competitive circumstances is almost certainly affected by the resources one brings to the conflict. If two 10-year-old children both share the goal of becoming 3rd grade teachers, which one is more likely to do so, all else equal: the one from a wealthy family or the one from a poor family? In contemporary societies, education and training are key components in office competition, and they both are expensive. Moreover, networking affects the selection of people for positions, and there can be little doubt that networking is a top-down phenomenon: the more successful, powerful, and (by correlation) wealthier the recommender, the more influential his recommendation. And who is more likely to be recommended by a successful, powerful, and rich mentor, all else equal: a rich person or a poor person? And is the average income of elected officials higher or lower than the

population average? Income helps a person gain access to all positions in society, and this will include offices that have some impact on the enculturation of youth.

Finally, even expertise is correlated with income. Becoming an expert on some important matter requires education and training, which, again, costs money. Or one can obtain expertise with life experience, but here again the wealthy have more resources to travel the globe and do the things that broaden one's understanding of the human condition. Achieving the reputation of being an expert on some topic almost certainly is affected by wealth.

Thus, all of the achievement mechanisms can be invoked to motivate the idea that people with higher incomes are more likely to enculture others. And, as was mentioned above, the central assumption here does not really involve income at all. The central idea is actually effort. In the model, people who do not work will not achieve status, fame, interactions, competitive offices, or expertise. As a result, they will have less influence on the tastes of the next generation. The item "y" that translates work into cultural influence is most intuitively thought of as income, but that is not necessary. More accurately, y is any kind of resource that a) is obtained through effort, b) is an element of a person's utility and well-being, but is not the sole element of either, and c) confers status, fame, interactions, competitive office, or expertise. Any number of things might fit these conditions, including income but also things like knowledge, mass communications, and power. Power, for example, requires effort, is directly enjoyed by those who have it, and confers status, fame, and competitive offices in virtually all human societies. In any case, the point is that those who have tastes for such things as knowledge, mass communication, power, and income will devote more effort to obtaining them than those

who do not have such tastes; those who obtain knowledge, mass communication, power, and income will also achieve status, fame, encounters, offices, and expertise; and those who have such achievements will have more influence on the tastes of the next generation than those who do not. The model simply reduces all of this to the assumptions that a) work produces income and b) income affects cultural transmission. This allows us to explore in a simple fashion the impact of cultural transmission on well-being.

Figure 2 shows the result of simulating the same model as above, with the addition now of an income effect in the adoption probability π. The two parameters q_y and q_h have been set to 4 and 3 respectively. As in Figure 1, population growth is zero at two points, s0 = 0 and s0 = 1, but now the only asymptotic equilibrium is at s0 = 0. The presence of the status mechanism makes the net conversion rate from ant to grasshopper negative: B = -0.02028. In each generation, there is a net outflow from the grasshopper population to the ant population, so that the only stable point involves zero grasshoppers. Everyone is an ant - and less happy than they would be if they were grasshoppers.

In the presence of an achievement mechanism, a taste for behavior that lowers well-being can flourish. "Achievement bias" occurs when an achievement mechanism is so strong that the resulting pattern of preferences produces a level of well-being that is below the maximum. One could quantify the achievement bias as the aggregate of deviations of each person's well-being (however measured) from his maximum. In this model, well-being is measured by the misery index, and each person, being an ant, is 0.19 points from his optimum. If there are 1,000 people in the society, the achievement bias produced by the introduction of the status mechanism would be (1,000)(0.19) = 190. One

assesses the value or harm of a cultural transmission mechanism by the amount of achievement bias it causes or removes.

The point of the simulation is that things like the status mechanism introduce a force that competes with well-being in influencing the evolution of preferences. If the competing force is strong, it may bias preferences away from activities that make people happy. The presence of achievement mechanisms is a necessary but not sufficient condition for achievement bias, however. If the status mechanism is weak (say, with q_y = 1 instead of 4), it may not dominate the effect of well-being.

This result does not depend on the insularity of society. Suppose that the ants and grasshoppers are insulated from one another to a different degree, so that the values of p and δ differ for the two groups. The net conversion rate is now expressed as

(10) $$B = p_a \delta_a \pi_g - p_g \delta_g \pi_a$$

In particular, suppose that adult ants are able to keep to themselves, so that naïve young grasshoppers only rarely encounter them. One might think of the harder-working and richer ants as being able to close off their world from less wealthy outsiders. As a result, the parameter δ_g will be lower than δ_a. Young ants go out into the world, see more of it, and hence have a higher chance of meeting people who are different; young grasshoppers go out into the world but remain locked within their own culture, and have less chance of meeting people who are different. This encourages conversions of ants to grasshoppers but discourages conversions in the other direction; it raises B. Figure 3 shows the effect: at all values of s0 between 0 and 1, B is still negative (we continue to assume q_y = 4 and q_h = 3) but now smaller in absolute value. It follows that differential degrees of social isolation can slow the rate of convergence to the unhappy equilibrium. Moreover, if they

were sufficiently strong, social isolation effects could swamp achievement effects. For example, perhaps young grasshoppers never even see ants, so $\delta_g = 0$, $B > 0$, and $s^* = 1$ as in Figure 1. But the sufficiency point remains: a status mechanism *may* cause unhappy tastes to flourish, under certain parameters, even if those who hold those tastes isolate themselves to a significant degree.

C. Mid-life corrections and natural selection

To this point, the model has ignored two important aspects of cultural reality. The first is the fact that the tastes one receives in youth may change in the course of life. Becker's (1996) work studies this phenomenon extensively. The second is the fact that choices affect health, and mortality affects the pool of people who can pass culture on to others.

To bring in these factors, let us add another stage to the life cycle. In the previous sections, the life cycle has two stages: one is born and is encultured, then one becomes an adult and encultures others. Now we assume that birth and enculturation is followed by a time period in which one does not enculture others. In this period, a person may devote time to modifying her own preferences. She may also die. Those who survive this period carry their preferences, which may or may not have changed, on to the third stage, during which they transmit their tastes to young people of the next generation.

The transmission of preferences follows the same rules as before, so that

(11) $$s_1 = s_0' + s_0'(1 - s_0')[p\delta\pi_g - p\delta\pi_a]$$

where s_0' is the fraction of grasshoppers in the population after mid-life conversion and mortality. This fraction is determined by the frequency of grasshoppers in generation 0:

(12) $$s_0' = s_0 + z_c(1 - s_0) - z_x s_0$$

where z_c is the probability that an ant will convert to grasshopper in mid-life, and z_x is the probability that a grasshopper will die before reaching the last stage of life. We assume that z_c is a function of the differential happiness of the two groups, while z_x is a function of the differential income of the two groups.

As for the first, the intuition is that well-being is eventually the dominant factor in the way people mold their own preferences. Given enough time, every person would be able to give themselves a utility function whose maximization also gave them the highest level of emotional satisfaction. Over time, people gradually learn which goals they have pursued in the past lead in fact to deeper well-being and which do not. Thus, mid-life corrections should typically be toward more happy states. In the context of this model, grasshoppers are always happier than ants, so we should allow some ants to switch to grasshopper in mid-life. (It is not necessary to allow conversions in the other direction, since they would not occur unless some criterion other than well-being were to cause them.)

As for differential mortality, the intuition comes from the fact that health status is strongly correlated with income. Medical care is a costlygood and access to it is often regulated by price. Even if it is not, however, mortality may differ because of behavioral effects, as poorer people have lower self-esteem and more destructive habits. Wealthier people have more access to health education and medical know-how. Even though income is often associated with hard work and stress, studies show that the net effect of income on health is beneficial (see Attanasio and Hoynes, 2000, and the references

therein). Thus, in the model, we would want to assume that some grasshoppers die early, and that this depends on their incomes.

We add these aspects to the model in the following way. First, we assume that mid-life conversions of ants to grasshoppers is determined by

(13) $$z_c = c[q_h(m_a - m_g)]$$

The term inside the brackets is the weight received by happiness in the conversion probability of young ants to grasshopper, and c is a parameter ($0 < c < 1$). The idea is that by mid-life the probability of conversion depends entirely on happiness, so that q_y is zero. However, in later life a person can be caught in a net of social obligations and norms, and no longer has the same freedom of thought and action as in youth. Moreover, the process of changing one's own tastes can be difficult and time consuming, so adults will be less likely, ceteris paribus, to seek change as they age. The conversion probability is therefore reduced by the factor c.

Second, the differential mortality of grasshoppers is given by

(14) $$z_x = x(y_a - y_g)$$

where x is a parameter. Here the idea is that relative deprivation has the strongest effect on health status (Wilkinson, 1996; Eibner, 2001). The mortality factor x is positive and may be smaller or larger than one, depending on the units of income. In the simulation here I initially choose x so that the conversion and mortality probabilities are about the same.

The main effect of adding an intermediate life stage is that the stable, zero growth points of the system are no longer at 0 and 1. The $\dot{s}$ function becomes

(15) $$\dot{s} = z_c(1-s_0) - z_x s_0 + (s_0 + z_c(1-s_0) - z_x s_0)(1-s_0 - z_c(1-s_0) + z_x s_0)B$$

This is a quadratic equation that does not reduce to $s_0(1-s_0)$.

Figure 4 shows how the addition of conversion and mortality affects the simulation. The conversion factor c = 0.2, while the mortality factor x = 0.5. As a result, the probability of mid-life conversion of ants to grasshopper is 11 percent, while the death rate of grasshoppers is 10 percent. The status mechanism is still active, with $q_y = 4$ as before. For the particular parameters in Figure 4, the zeroes of (17) are 0.51 and 17.66. The first of these can be seen in the figure; the $\dot{s}$ function is positive when s_0 is less than 0.51, and negative when s_0 is above 0.51; the frequency of grasshoppers grows when it is below 0.51 and falls when it is above that level. Thus, $s^* = 0.51$ is an asymptotic equilibrium of the system.

Relative to Figure 2, we see that adding differential mortality and mid-life conversions shifts the equilibrium from s = 0 to s = 0.51; not all grasshoppers die out, despite the bias introduced by the status mechanism. On the other hand, the unhappy ants do not die out either. Instead, with the added realism in the model, the population comes to rest with an even mix of the two types. At this equilibrium, 49 percent of the population has a misery index of 0.19; for a population of 1,000, the measure of achievement bias would be 93.1. Achievement bias exists even when one accounts for mid-life corrections and natural selection.

Figure 5 shows the impact of a change in the status mechanism under these more realistic conditions. We have reduced the status mechanism's power by lowering q_y from 4 to 1. The equilibrium shifts upward to $s^* = 0.57$. Now only 43 percent of the population consists of unhappy ants, and the metric of achievement bias falls from 93.1 to 81.7. In

other words, the system responds in plausible ways to changes in cultural transmission mechanisms: if you reduce the influence of the status mechanism, cultural evolution will increase the number of happy people.

This simulated society responds plausibly to variations in the other parameters. If we make grasshoppers more likely to be encultured by their parents (raising p_g), the equilibrium shifts upward; if we make the same assumption for ants, it shifts downward. If the differential mortality of grasshoppers rises, the equilibrium involves fewer grasshoppers; if more mid-life ants convert to grasshopper, the equilibrium involves fewer ants. Changes in the overall permeability of society (p_a, p_g, δ_a, δ_g) do not change the equilibrium point (so long as both groups are equally permeable), but do change the rate at which society converges to it. As one alters parameters, it is of course possible to produce virtually any mix of ants and grasshoppers as a social equilibrium. This only strengthens the basic point, however, which is that it is not necessarily the case that cultural evolution always eliminates preferences that lower well-being. Rather, it is easy, in almost any parameter set, to find an alteration which results in an equilibrium with lower well-being.

Other than this, the wage parameter deserves closer attention. Over time, technological progress and economic growth could raise the wage rate. Figure 6 shows the effect on tastes. It uses the parameters of Figure 4, except that the wage of 0.4 is doubled to 0.8. The equilibrium frequency of grasshoppers falls from 0.51 to 0.35. As wages rise, the opportunity cost of leisure rises. Ants and grasshoppers both work more, but ants, with their greater utility of income, do so disproportionately. Even though the difference in tastes between the two groups has not changed, the income gap between

them widens. As a result, the impact of the status mechanism increases: the increased gap in income also widens the differences in status, and makes the higher status of the ants more apparent and more worthy of emulation. Moreover, the income gap makes ants relatively still more effective than grasshoppers at surviving to the third stage of life, so that the differential mortality of grasshoppers rises. Finally, the widening income gap makes ants more unhappy than they were (we continue to assume that grasshoppers are at the bliss point), which makes mid life conversions more likely.[21] The net effect of these changes is to lower the frequency of grasshoppers. Thus, one can construct a model with plausible parameters and mechanisms in which technological progress increases the number of unhappy people in society. Increases in per capita income can be consistent with lower well-being. Again, this is not an argument that such an outcome is necessary, only that it is possible.

D. Group selection

Within a given society, then, cultural transmission systems do not necessarily select only for tastes that raise well-being. Different societies may have different parameters and therefore different combinations of happiness, income, and leisure; in some cultures, everyone may be a grasshopper, in others everyone is an ant; in still others there may be a mix of types.

Suppose the world consists of a very large number of separate societies. There is no migration across them.[22] Each society is endowed with unique parameters governing its cultural transmission processes, its economic development, and so on. As a result, each society i has a population frequency of grasshoppers s_i.

Now suppose that the societies are subject to evolutionary pressures themselves. Perhaps resources are so limited that some of the groups run out of essential goods and die off. Perhaps they engage in war. To model group selection in the context of the income and leisure model, we have to make some assumptions about the way that the selection mechanism chooses which groups survive.

Generally speaking, the societies in this model are characterized by different mixes of income, leisure, and happiness. Given this, the most plausible selection mechanism would seem to be that societies with more income are more likely to survive than others.[23] If survival depends on the possession of material goods, then it is income, and not leisure or happiness, that would determine fitness. In the fable, the grasshoppers get hungry and weak when the winter comes; perhaps they die. And if survival depends on war, only income matters. In the fable, the industrious ant shares some of his abundant resources with the weakened grasshopper, but that is why it is a fable. If the ants acted like real humans instead of imaginary insects, they would more likely wait for the grasshoppers to weaken and then push them off their land, killing most and leaving the rest to starve. If group selection operates at all, it seems most likely to operate on incomes rather than leisure or well-being.

We will assume that the probability that a group survives from one period to the next is given by $\Pr(\text{survival} \mid y_i) = \rho y_i$, where y_i is the per capita income of the society in question, and $\rho > 0$. Suppose there are two types of societies in the world, type j and type k, with equilibrium grasshopper frequencies of s_j and s_k respectively. Let the j type society have more grasshoppers, hence $s_j > s_k$. If wage levels in the two societies are sufficiently close, this will imply that $y_j < y_k$.[24] When extinction occurs, new societies are

formed as colonies from the existing societies, in proportion to the current mix of those societies. Thus, if f_0 is the frequency of type j societies in period 0, then the probability that an extinct society is replaced by a new society of type j is f_0, and the probability it is replaced by a society of type k is $(1-f_0)$.

The dynamics of the system are given by

(16) $$f_1 = \rho y_j f_0 + (1-\rho y_j) f_0^2 + (1-\rho y_k)(1-f_0) f_0$$

The first term is the number of type j societies that survive, the second is the number of type j societies that become extinct but are replaced by another type j society, and the third is the number of type k societies that become extinct and are replaced by a type j society. The dynamics reduce to

(17) $$\dot{f} = f_0(1-f_0)(\rho y_j - \rho y_k)$$

The term $\rho y_j - \rho y_k$ is negative, which means that the asymptotic equilibrium of the system is $f^* = 0$. Group selection annihilates grasshoppers.

It would be no less difficult to build simple models of group selection in which some grasshopper societies would survive. Such models would have to assume that group selection operates at least as strongly on leisure and well-being as on income. At the level of individual taste selection, such assumptions make a great deal of sense. Indeed the societal model in the previous sections based the survival of tastes on a complex mix of income, leisure, and well-being. At the level of whole societies, however, the case is harder to make. At times, the leaders of one society may have envied the leisure and well-being of other societies, and may have tried to emulate those outcomes in their own

society. But the envy of wealth and the respect earned by superior armies have probably been the more powerful force in human history.

Ammerman and Cavalli-Sforza (1971, p. 685) present a diagram showing the spread of early farming, from a presumed origin near Jericho, northwestwards to Ireland. The picture tells the story that between 6000 and 3000 BCE, farm societies gradually but inexorably replaced hunter-gatherer societies, one by one, from one end of Europe to the other. They did so probably not because farming generally makes a person happier, but because farms produce a great deal of food with relatively low risk.

Indeed, it seems unlikely that group selection processes would encourage the formation of grasshopper societies if within-group pressures were more likely to produce ants. If anything, group selection would further encourage the growth of ant societies. The possibility argument of the preceding sections seems largely immune to group selection processes. Even with group selection, it is still possible for tastes for immiserating behavior to persist in cultural equilibrium.

VI. Implications: Progress, Civilization, Misery

The main point of the paper is that preferences for actions which lead to relatively low well-being may persist in cultural equilibrium. Culture may give us goals whose pursuit will make us unhappy. This happens because of certain mechanisms of cultural transmission, related to such goals as income, power, knowledge, and mass communication. Pursuit of these goals gives a person a number of important social achievements: status, fame, offices, interactions, and expertise. People with such achievements broadcast their goals more loudly, and thereby amplify the importance of

these goals to the young. Such goals propagate in society to the disadvantage of other goals. Societies populated by people with such goals generally have more wealth, more power, superior knowledge, and better mass communications. They dominate other societies and are envied by them. At both the societal and individual level, tastes for income, power, knowledge, and mass communication have an evolutionary advantage over other tastes.

Great achievements build great civilizations, but they contribute only partially to human well-being. Happiness may require such things, but it also requires other things that do not enjoy similar advantages in cultural evolution. It requires good relations with an intimate partner, as well as with one's children, parents, siblings, and friends; it requires a sense of meaningful existence in the cosmos; it requires inner peace. Perhaps there are cultural mechanisms that broadcast tastes for these things as loudly as the achievement mechanism broadcasts tastes for wealth and power.[25] If so, then humans would be developing not only ever-increasing wealth, power, knowledge, and mass communications, but also ever-increasingly good relations with their families, their gods, and themselves. This, however, seems not to be the case.

References

Ammerman, A. and L. Cavalli-Sforza (1971), "Measuring the Rate of Spread of Early Farming in Europe," *Man* 6, 674-88.

Argyle, Michael (1999), "Causes and Correlates of Happiness," in Daniel Kahneman, Ed Diener, and Norbert Schwarz (eds.), *Well-Being: The Foundations of Hedonic Psychology*, Russell Sage, 353-373.

Attanasio, orazio P. and Hilary W. Hoynes (2000), "Differential Mortality and Wealth Accumulation," *Journal of Human Resources* 35(1), 1-29.

Banerjee, A. V. (1992), "A Simple Model of Herd Behavior," *Quarterly Journal of Economics* 107, 797-817.

Barkow, J., L. Cosmidies, and J. Tooby (1992), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, New York: Oxford University Press.

Becker, Gary S. and Kevin M. Murphy (1988), "A Theory of Rational Addiction," *Journal of Political Economy* 96(4), 675-700.

Becker, Gary S. (1996), *Accounting for Tastes*, Harvard University Press.

Ben-Ner, Avner, and Louis Putterman (1998), "Values and Institutions in Economic Analysis," in A. Ben-Ner and L. Putterman (eds.), *Economics, Values, and Organization*, Cambridge University Press.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* 100, 992-1026.

Bisin, Alberto and Thierry Verdier (2000), "'Beyond the Melting Pot': Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics* 115(3), 955-88.

Blanchflower, D. G. and A. J. Oswald (2000), "Well-Being Over Time In Britain and the USA," NBER Working Paper No. 7487, National Bureau of Economic Research.

Bloom, Harold (1998), *Shakespeare: The Invention of the Human*, Riverhead Books.

Bowles, Samuel (1998), "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions," *Journal of Economic Literature* 36(1), 75-111.

Bowles, Samuel and Herbert Gintis (1998), "The Moral Economy of Communities: Structured Populations and the Evolution of Pro-Social Norms," *Evolution and Human Behavior* 19, 3-25.

Boyd, Robert and Peter J. Richerson (1985), *Culture and the Evolutionary Process*, University of Chicago Press.

_____ (1994), "The Evolution of Norms: An Anthropological View," *Journal of Institutional and Theoretical Economics* 150(1), 72-87.

Brickman, P., D. Coates and R. Janoff-Bulman (1978), "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology* 37, 917-27.

Cavalli-Sforza, L. L. and M. W. Feldman (1981), *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton University Press.

Cowen, Tyler and Alexander Tabarrok (2000), "An Economic Theory of Avant-Garde and Popular Art, or High and Low Culture," *Southern Economic* Journal 67(2), 232-53.

Dawkins, Richard (1976), *The Selfish Gene*, Oxford University Press.

Easterlin, Richard A. (1974), "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," in P. A. David and M. W. Reder (eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramowitz*, Academic Press, 89-125.

_____ (1995), "Will Raising the Incomes of All Increase the Happiness of All?" *Journal of Economic Behavior and Organization*, 27(1) 35-48.

Eibner, Christine E. (2001), "Income, Relative Deprivation, and Mortality: Evidence From Individual-Level Data," University of Maryland, mimeo.

Frank, Robert H. (1987), "If *Homo Economicus* Could Choose His Own Utility Function, Would He Want One With a Conscience?" *American Economic Review* 77(4) 593-604.

Frey, Bruno S. and Alois Stutzer (2000), "Happiness, Economy and Institutions," *Economic Journal* 110, 918-38.

Gintis, Herbert (1972), "A Radical Analysis of Welfare Economics and Individual Development," *Quarterly Journal of Economics* 86(4), 572-99.

_____ (2000), *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction*, Princeton University Press.

Greif, Avner (1994), "Cultural Beliefs and the Organization of Society: Historical and Theoretical Reflection on Collectivist and Individualist Societies," *Journal of Political Economy* 102(5), 912-50.

Gross, David and Nicholas S. Souleles (2000), "An Empirical Analysis of Personal Bankruptcy and Delinquency," Wharton, mimeo.

Guth, Werner and Axel Ockenfels (2000), "Evolutionary Norm Enforcement," *Journal of Institutional and Theoretical Economics* 156, 335-347.

_____ and M. Yaari (1992), "An Evolutionary Approach to Explain Reciprocal Behavior in a Simple Strategic Game," in Ulrich Witt (ed.), *Explaining Process and Change – Approaches to Evolutionary Economics*, University of Michigan Press, 23-34.

Harrington, J. E. (1999), "Rigidity in Social Systems," *Journal of Political Economy* 107(1), 40-64.

Hausman, Daniel M. and Michael S. McPherson (1996), *Economic Analysis and Moral Philosophy*, Cambridge University Press.

Kahneman, Daniel, Ed Diener, and Norbert Schwarz (eds.) (1999), *Well-Being: The Foundations of Hedonic Psychology*, Russell Sage.

Kuran, Timur (1991), "Cognitive Limitations and Preference Evolution," *Journal of Institutional and Theoretical Economics* 147, 241-73.

Ledoux, Joseph (1996), *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, NY: Simon and Schuster.

Lumsden, Charles J. and Edward O. Wilson (1981), *Genes, Mind and Culture: The Coevolutionary Process*, Harvard University Press.

Miller, Geoffrey (2000), *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*, NY: Doubleday.

Nozick, Robert (1974), *Anarchy, State and Utopia*, Basic Books.

Putnam, Robert D. (1995), "Bowling Alone: America's Declining Social Capital," *Journal of Democracy* 6(1), 65-78.

Rogers, Alan R. (1994), "Evolution of Time Preference by Natural Selection," *American Economic Review* 84(3), 460-81.

Ruyle, Eugene E. (1973), "Genetic and Cultural Pools: Some Suggestions for a Unified Theory of Biocultural Evolution," *Human Ecology* 1(3), 202-215.

Schelling, Thomas C. (1984), "Self-Command in Practice, in Policy, and in a Theory of Rational Choice," *American Economic Review* 74, 1-11.

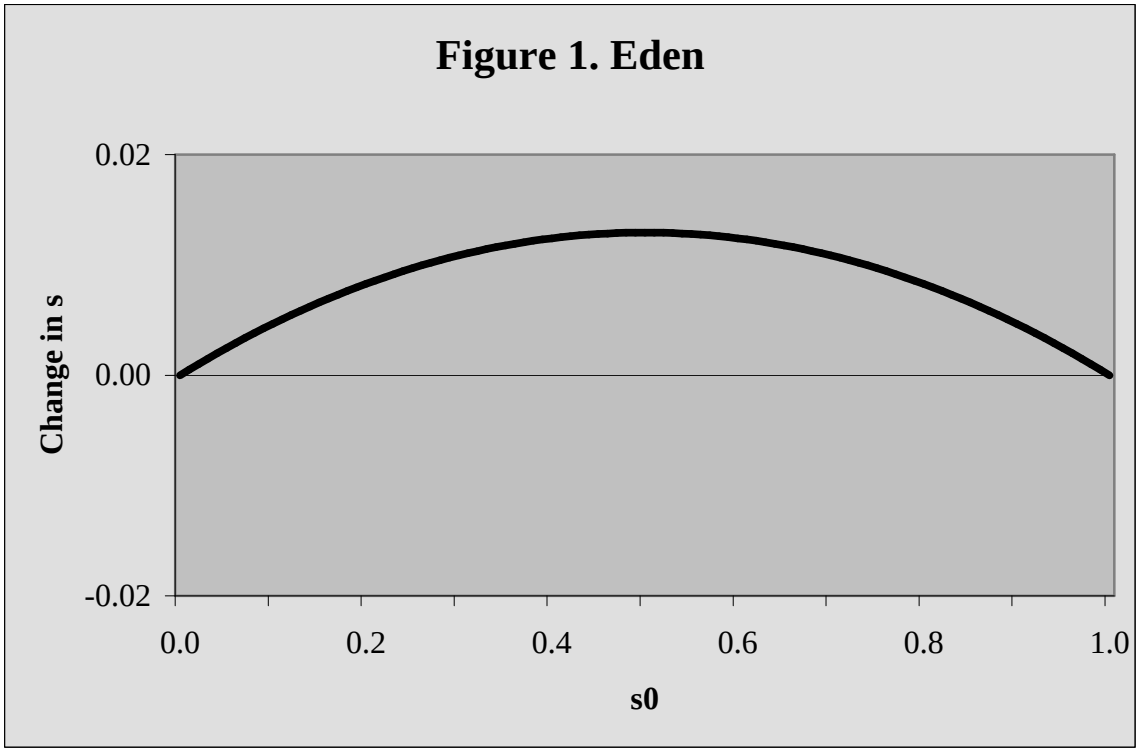
Sen, Amartya (1993), "Capability and Well-Being," in Martha Nussbaum and Amartya Sen, eds., *The Quality of Life*, Clarendon, 30-53.

Thaler, Richard and H. M. Sheffrin (1981), "An Economic Theory of Self-Control," *Journal of Political Economy* 89(2), 392-406.

Weibull, Jőrgen (1995), *Evolutionary Game Theory*, MIT Press.

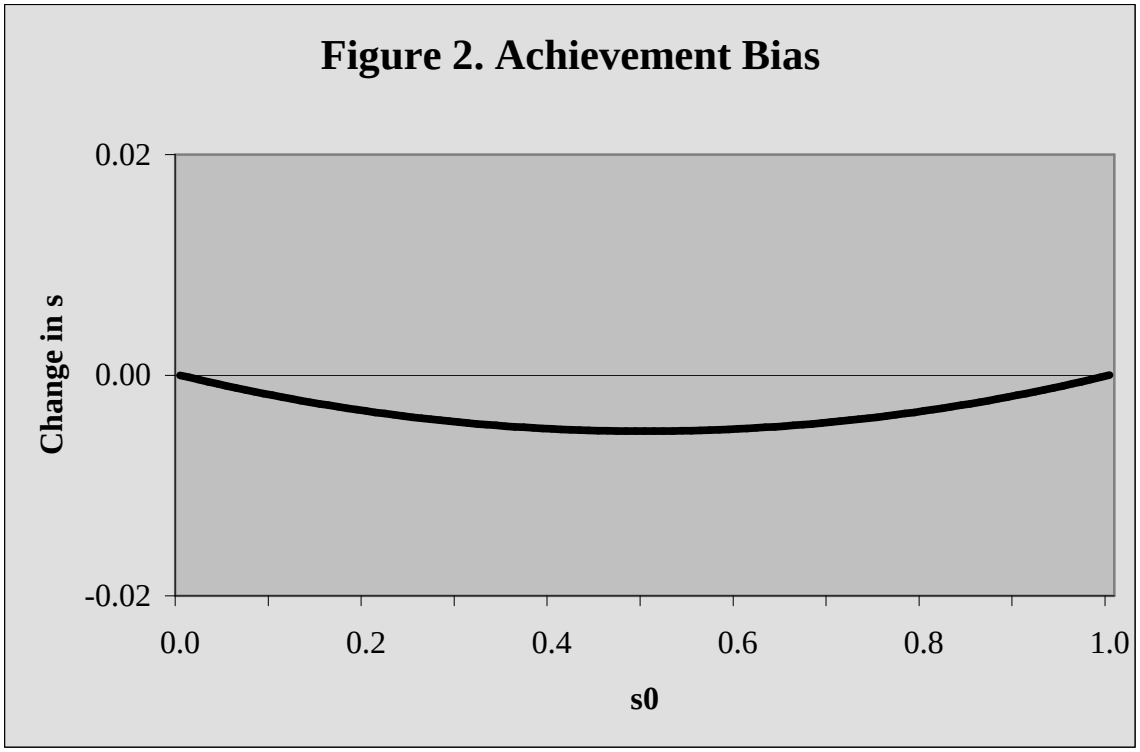
Wilkinson, Richard G. (1996), *Unhealthy Societies: The Afflictions of Inequality*, Routledge.

Wilson, William Julius (1987), *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, University of Chicago Press.



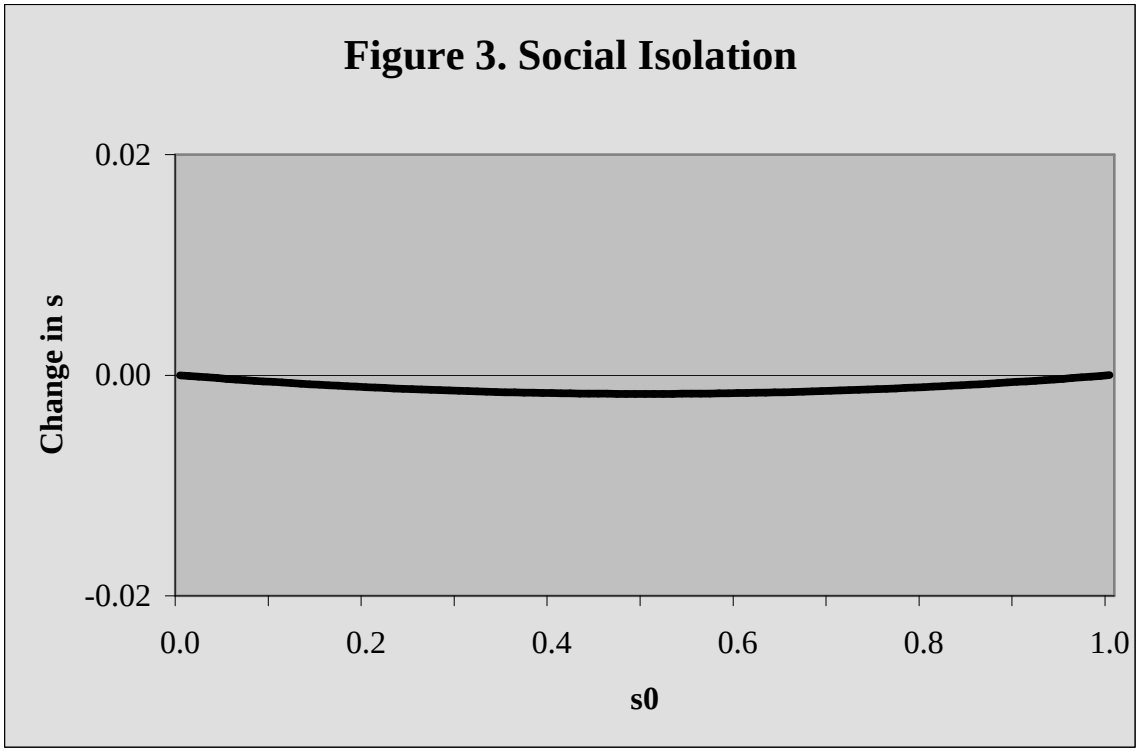
Parameter Set 1:

p_a	0.3	c	0.00
δ_a	0.3	z_c	0.00
p_g	0.3	x	0.0
δ_g	0.3	z_x	0.00
α_g	0.25	y_0	0.1
α_a	0.75	L_0	0.75
T	1	m_a	0.19
w	0.4	m_g	0
y_a	0.3	q_y	0
y_g	0.1	q_h	3
L_a	0.25	π_a	0.00
L_g	0.75	π_g	0.57
B	0.05172	X-intercepts of $\dot{s}(s_0)$:	1.00
			0.00



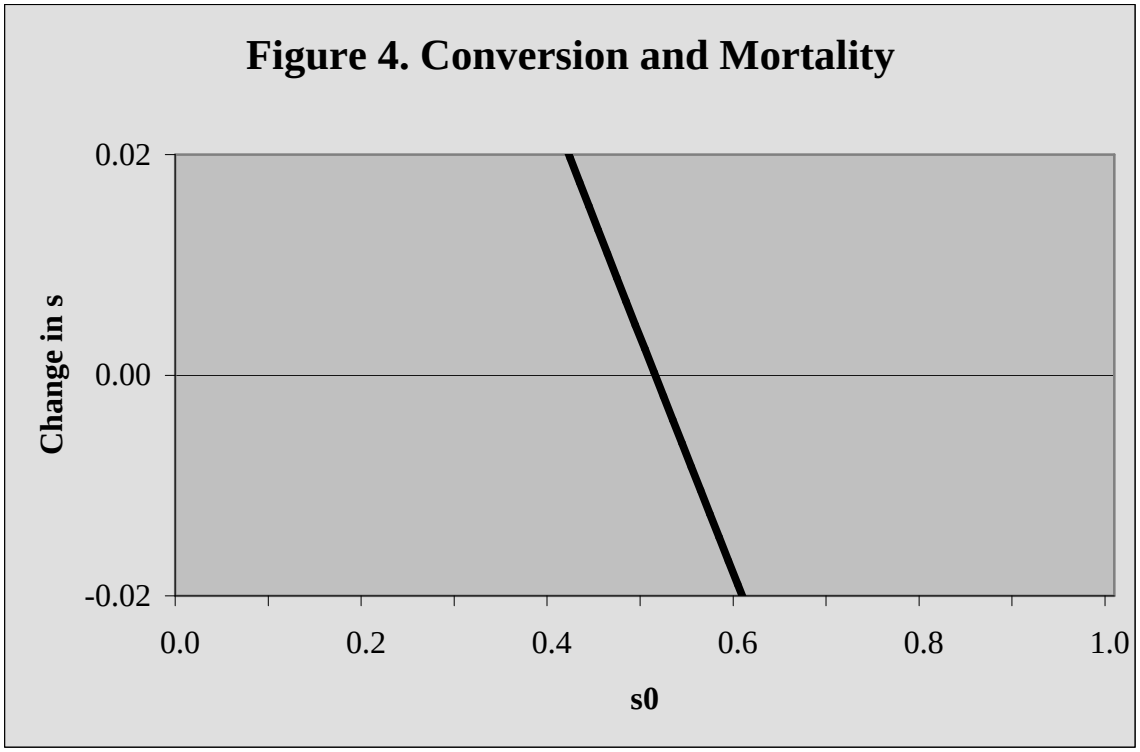
Parameter Set 2:

p_a	0.3	c	0.00
δ_a	0.3	z_c	0.00
p_g	0.3	x	0.0
δ_g	0.3	z_x	0.00
α_g	0.25	y_0	0.1
α_a	0.75	L_0	0.75
T	1	m_a	0.19
w	0.4	m_g	0
y_a	0.3	q_y	4
y_g	0.1	q_h	3
L_a	0.25	π_a	0.23
L_g	0.75	π_g	0.00
B	-0.02028	X-intercepts of $\dot{s}(s_0)$:	1.00
			0.00



Parameter Set 3:

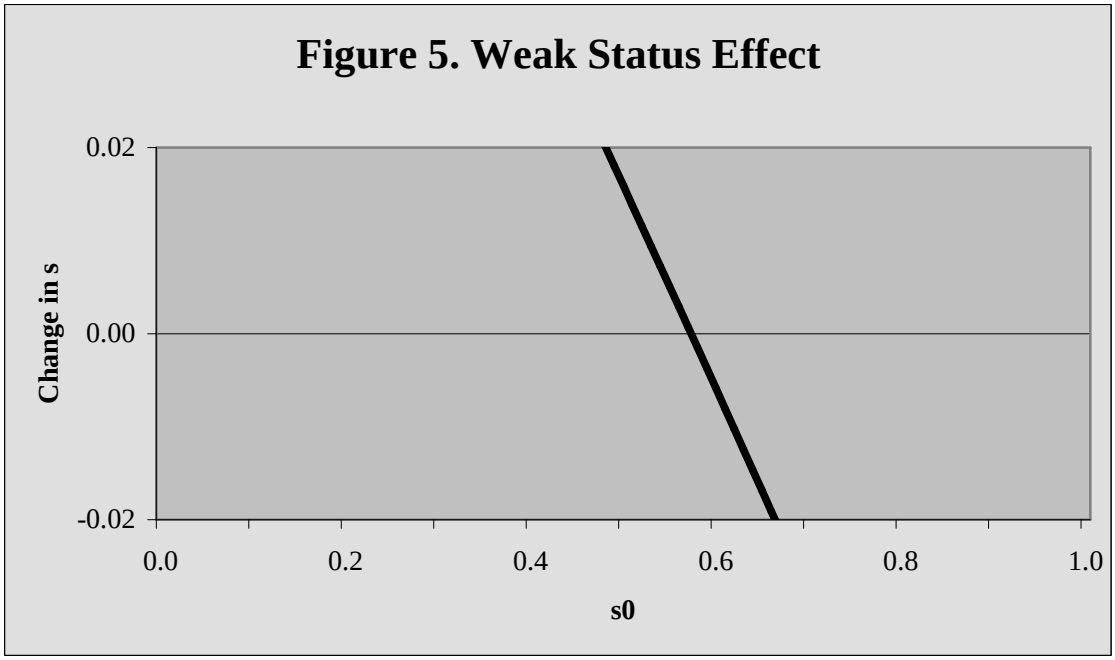
p_a	0.3	c	0.00
δ_a	0.3	z_c	0.00
p_g	0.3	x	0.0
δ_g	0.1	z_x	0.00
α_g	0.25	y_0	0.1
α_a	0.75	L_0	0.75
T	1	m_a	0.19
w	0.4	m_g	0
y_a	0.3	q_y	4
y_g	0.1	q_h	3
L_a	0.25	π_a	0.23
L_g	0.75	π_g	0.00
B	-0.00676	X-intercepts of $\dot{s}(s_0)$:	1.00
			0.00



Parameter Set 4:

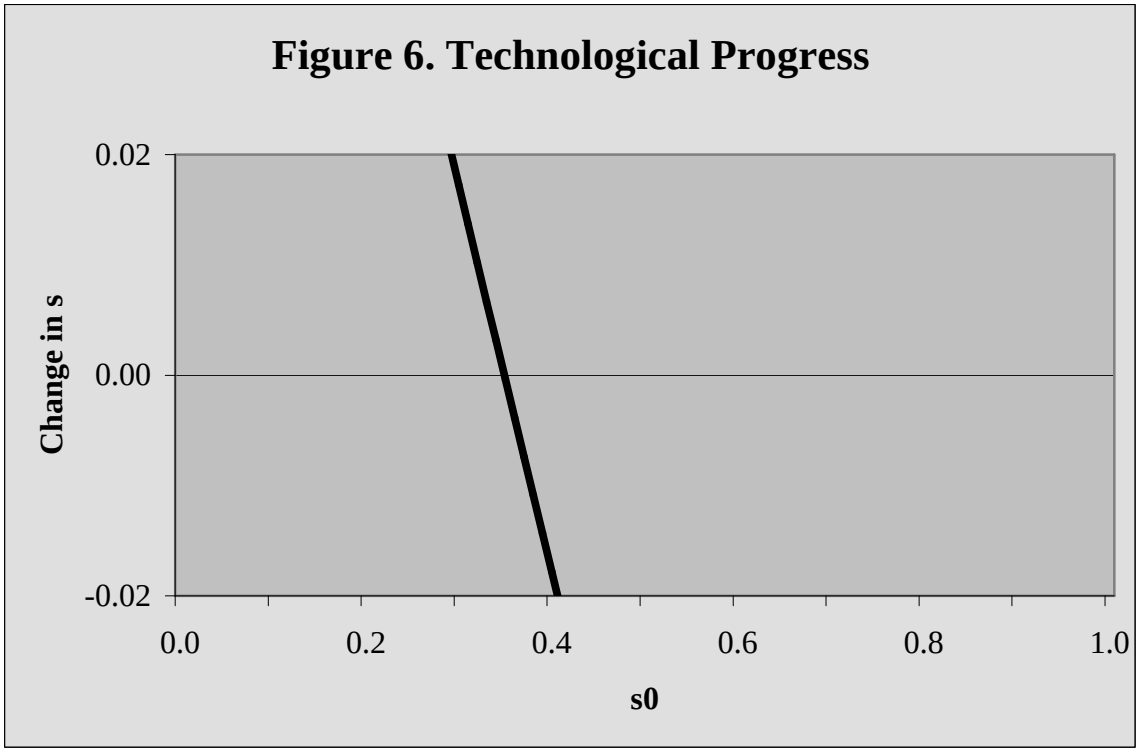
p_a	0.3	c	0.20
δ_a	0.3	z_c	0.11
p_g	0.3	x	0.50
δ_g	0.3	z_x	0.10
α_g	0.25	y_0	0.1
α_a	0.75	L_0	0.75
T	1	m_a	0.19
w	0.4	m_g	0
y_a	0.3	q_y	4
y_g	0.1	q_h	3
L_a	0.25	π_a	0.23
L_g	0.75	π_g	0.00
B	-0.02028	X-intercepts of $\dot{s}(s_0)$:	0.51
			17.66

Figure 5. Weak Status Effect



Parameter Set 5:

p_a	0.3	c	0.20
δ_a	0.3	z_c	0.11
p_g	0.3	x	0.50
δ_g	0.3	z_x	0.10
α_g	0.25	y_0	0.1
α_a	0.75	L_0	0.75
T	1	m_a	0.19
w	0.4	m_g	0
y_a	0.3	q_y	1
y_g	0.1	q_h	3
L_a	0.25	π_a	0.00
L_g	0.75	π_g	0.37
B	0.03372	X-intercepts of $\dot{s}(s_0)$:	0.57
			-9.93



Parameter Set 6:

p_a	0.3	c	0.20
δ_a	0.3	z_c	0.14
p_g	0.3	x	0.50
δ_g	0.3	z_x	0.20
α_g	0.25	y_0	0.2
α_a	0.75	L_0	0.75
T	1	m_a	0.23
w	0.8	m_g	0
y_a	0.6	q_y	4
y_g	0.2	q_h	3
L_a	0.25	π_a	0.92
L_g	0.75	π_g	0.00
B	-0.08250	X-intercepts of $\dot{s}(s_0)$:	0.35
			10.02

[1] The initial preferences are important, since they are not easy to change. In Becker's (1996) approach, one's preferences can only be changed through the accumulation of certain actions, which takes time. One of Shakespeare's most important contributions to human thought is the idea that self-modification only comes through self-understanding, and can be very difficult, time-consuming, and costly (Bloom, 1998). Finally, economics requires that the agent cannot immediately change her own preferences, otherwise a preference-based choice model would make no sense. For all of these reasons, the initial preferences at the time of endowment, at the end of childhood, deserve careful study.

[2] Sahlins (1976) argues that culture shapes preferences, and therefore culture must be the start of a choice-based theory of human behavior. He does not, however, propose any theory by which choices affect culture, as they certainly do.

[3] Outside of economics there is a large literature devoted to the evolution of cultural traits in general, but none of it focuses specifically on traits that economists would identify as preferences. In anthropology, not much distinction is made between a trait that gives the agent a taste for some behavior, as opposed to a trait that gives the agent both the taste for the behavior and also the resources necessary to engage in it. The trait tends to be identified at the level of behavior; one is either a 'hell-raiser' or not (Boyd and Richerson, 1985). This paper focuses instead on the agent's tastes for hell-raising, regardless of whether they in fact lead to rowdy behavior.

[4] True, power-seeking will maximize the utility of a person emerging from childhood, but utility maximization and happiness maximization are not the same thing. Thus, it may be the case that a person would devote her life to the pursuit of power, and eventually become leader, and yet find herself unhappy even though her utility is at its maximum. By definition, utility always guides behavior; happiness often does not, as is apparent from the flourishing of the therapeutic professions. Indeed one could view the objective of therapy (and maturation more generally) as an ever-greater consistency between the utility function and the "happiness function." Many authors have argued that utility and human well-being are not equivalent; their arguments are critical for this paper and will be discussed in detail below.

[5] The process of molding the self takes up considerable social resources, including both time and money. Becker's (1996) approach to endogenous preferences is to assume that the current utility of an action depends on past choices. An agent could give herself a specific current utility function by accumulating choices accordingly. The process would take time, and, like any capital accumulation process, would require sacrificing goals of the moment. Thus, if changing tastes is costly, the time path of preferences through the life cycle will depend on initial conditions. Moreover, if we adopt a more psychotherapeutic, Shakespearian view of preference change, we will have to admit that immediate convergence to a desired utility function may not be possible. The mind is not perfectly and immediately mutable. People do things that will make them unhappy. They continue to do them long after becoming aware of the connection. Our initial tastes can have very long shadows.

[6] Well-being and happiness are not necessarily the same, but the distinction between them does not affect anything in the argument here. The point is that both are distinct from utility.

[7] Throughout the paper I will refer to fitness under natural selection (i.e. the fact that human beings who are more poorly adapted to their environment are more likely to die before bearing children) as 'darwinian fitness' or 'biological fitness.'

[8] Bisin and Verdier's (2000) paper is a recent example. The population of agents is divided into types, the agents play games, and the agents with the highest payoffs are more fit and pass their type on to larger numbers of the succeeding generation. This leads to a dynamics in which the change in the frequency of the type depends on the type's current payoffs. In one mathematical form, this is called the replicator dynamics.

[9] Gintis notes that this is similar to meiotic drive, a force in biological evolution that allows organisms to retain attributes that are actually damaging to them.

[10] I will use the terms 'happiness' and 'well-being' interchangeably throughout. Nozick (1974) argues against the idea that human well-being is equivalent to a happy mental state, otherwise it would be acceptable to live life in a drugged state while hooked to an experience machine that produced nothing but good feelings in the brain. Such counter-examples notwithstanding, I will assume that practically speaking, well-being is always enhanced by increases in subjective human happiness, and that true human happiness is the sine qua non in human well-being.

[11] Lottery winners are a common example: surveyed before and after winning, they typically exhibit modest increases in wealth but little or no increase in emotional satisfaction or well-being (Brickman, Coates, and Janoff-Bulman, 1978). Marriage has a much larger impact on happiness than income (Argyle, 1999), yet the utility functions of many people, perhaps most people, and especially young people, are directed more toward career than relationships. True, young people do build relationships and learn about them through trial and error. And they also pick up career skills through odd jobs. While such learning might be reasonably successful, much of it is bound to be haphazard. It is interesting to note that people find such unstructured learning utterly unacceptable when it comes to careers, but not when it comes to relationships. They supplement their haphazard career learning with formal career training, but they do not supplement their haphazard relationship learning with formal relationship training. If they did, the formal education system would look quite different from the way it does at this writing. The formal education system in contemporary societies seems largely devoted to career and workplace preparation; aside from the occasional Human Sexuality class, most coursework is intended to improve skills that are either directly vocational (Drawing I, II, III for the fine arts or Accounting for pre-business) or involve general life preparation (Algebra; Plato). Relative to these, how much time is spent learning and honing the specific and well-known skills that intimate physical relationships require: communication, trust, openness, reliability, and above all, self-awareness? Could we not conceive of some kind of formal training, equivalent to the decades-long process of gaining familiarity with the techniques of language and mathematics, that would give people more familiarity with the techniques of self-assessment? Such education could exist, but it does not. Rather, young people devote years and years to formal schoolwork that prepares them almost exclusively for work. Since they do this largely voluntarily, and continue well into their adult years, it follows that their utility functions at emergence from childhood must be maximized at bundles that contain excellent careers and mediocre relationships. Unfortunately, studies of subjective well-being, as well as mature intuition, suggest that happiness is generally not maximized when the career is good and the relationship is not so good.

[12] Gross and Souleles (2000) report that people often keep low-interest liquid assets *and* high-interest credit card debts at the same time. This violates simple precepts of financial rationality. They suggest that self-control issues are the most likely explanation.

[13] Nozick (1974) argues against such a view of substantive well-being. It could be satisfied by hooking people up to experience machines that would make them happy. Sen (1993) and many others argue that human well-being depends more on the kind of life a person lives rather than an emotional state. In this paper I will assume that whatever happiness people obtain is acquired by the living of a good life rather than an experience machine.

[14] Ruyle (1973) argues that biological evolution has created our desire for emotional satisfaction, and this is how it has structured our behavior.

[15] An example that hits closer to home is the influence of the tenure process. Graduate students are typically a random draw from the population as far as an interest in being tenured goes; they may or may not care about it at first. Still, those who do care about it are more likely to be tenured in the end. As a result, the average tenured faculty member believes tenure to be a more important thing than does the average first-year graduate student. If senior faculty mentoring has any influence on graduate students, it will, on average, induce them to care more about tenure than they otherwise would.

[16] One could make an argument that achievement mechanisms may exist simply because culture exists. The capacity for culture involves the capacity to learn from others, which implies that there must be a capacity to teach others. Suppose, then, that culture exists if and only if all members of the population have a basic desire to teach others what they know. Thus, each member of the population has a basic urge to propagate her own tastes in the next generation. It follows that the famous will want children to imitate those who are famous, that the prestigious will want children to imitate those who have prestige, that the knowledgeable will want children to imitate those who have knowledge, and so on. Similarly, those who are unknown will want children to imitate the obscure, those who are humble will want children to imitate the humble, and those who are ignorant will want children to imitate those who are ignorant. However, the famous, prestigious, and knowledgeable people will generally enculture more children than will the obscure, humble, and ignorant people, simply because of the nature of fame, prestige, and knowledge in human society. Hence the achievement mechanisms may exist simply because ego is a basic human drive: all people want to be imitated, but the famous, the prestigious, and the knowledgeable have more power to convince naïve agents that it is *they*, and not some others, who should be imitated.

[17] I use these terms to avoid the cumbersome jargon of 'j-type agents' and 'k-type agents'.
[18] Assortive mating would be redundant, since below it is assumed that children are encultured mostly within their group anyway.
[19] In the simulation of the model, the values of this probability will be bounded at 0 and 1.
[20] We are assuming that preferences are transmitted via judgments of happiness and income. Both of these can be observed (albeit with error) and therefore compared across individuals. Utility cannot be observed and cannot be compared across individuals; utility cannot be the basis of theories of cultural preference transmission. Yet in many circumstances (e.g. bargaining theory), it makes sense to think of utility and income as equivalent, and in other cases (tax policy analysis), it makes sense to think of utility and well-being as equivalent. Thus one can see why payoff-based models are a plausible and intuitive initial approach to modeling cultural preference transmission. Since we often think of happiness and income as the payoffs of our actions, it makes sense to use payoffs as the standard for the transmission of tastes – even if, in reality, it is the happiness and the income, and not the payoffs themselves, which structure our decisions to adopt a taste or not. In this model, we examine cases where happiness and income are not necessarily the payoffs of our actions; they remain the standard of taste adoption, but not the standard of behavior.
[21] If we assumed that the original bundle (y_0, L_0) were still the bliss point, increases in wages would make even the grasshoppers less happy.
[22] In fact we will assume that there is no migration between societies. If there were, it would have effects similar to the cross-group enculturation and conversion processes within a society, which have already been described. By ruling these out, we effectively define a society as a group whose children can only be encultured by adult group members, and whose adults can only enculture the group's children.
[23] Greif (1994) argues that cultures are strongly path-dependent, and that the transition from an inferior cultural form to a superior one may take a great deal of time or perhaps may never happen. It is interesting that his study of Mediterranean traders is trying to explain why some cultures did *not* adopt a cultural institution that increased incomes. In other words, the failure of competitive selection pressures to convert a group to a higher-income cultural practice is treated as an anomaly, worthy of special modeling and explanation. Implicitly, the general rule must be that societies with higher incomes are copied by societies with lower incomes whenever social institutions are sufficiently flexible.
[24] If wages in type j societies were higher than those in type k societies, and if α_j and α_k were close to one another, it could be the case that a society with more grasshoppers would have higher average incomes. One could make the argument that a society which focuses more on well-being might be intellectually more vibrant and hence would have a higher level of development, hence higher wages. However, the within-society model above made the counter-argument that development itself, by increasing the wage, would reduce the number of grasshoppers. Ultimately, development is a matter of investment, which depends on savings, which in turn depends on income. In the fable, it is the ants and not the grasshoppers who have savings.
[25] It would not be difficult to build a model in which people sacrificed leisure to devote time to strengthening family relations. People with tastes for family-building would be happier, and so the selection of tastes based on well-being would favor tastes for family building. Also, people with tastes for family building would produce children who were less likely to commit suicide, lending an advantage in terms of natural selection. Undoubtedly such mechanisms exist and they explain the long-run persistence of the family to this point. I have chosen to focus on achievement mechanisms instead, however, because history seems to show them to be stronger. The objects that achievement mechanisms favor have grown more or less steadily throughout recorded time, and individuals in contemporary post-industrial civilizations enjoy the highest levels of income, power (in the form of freedoms), knowledge, and mass communications that have ever been experienced in human history. The objects that family mechanisms (and other relationship mechanisms, such as to gods and the self) favor have not grown steadily throughout recorded time. Indeed, one could argue that people in contemporary post-industrial civilizations suffer under the weakest relationships with intimate partners, families, gods, and the self that have ever been experienced in human history.