7

A Typology of Human Morality

Herbert Gintis

Abstract

This chapter suggests a typology of human morality based on gene–culture coevolution, the rational actor model, and behavioral game theory. The basic principles are that human morality is the product of an evolutionary dynamic in which evolving culture makes new behaviors fitness enhancing, thus altering our genetic constitution. It is thus predicated upon an evolved set of human genetic predispositions and consists of the capacity to conceptualize and value a moral realm governing behavior beyond consequentialist reasoning.

Introduction

By behavioral morality, I mean the set of moral rules that govern the choices that people make. Behavioral morality is distinct from the normative morality to which philosophers and theologians propose that people ought to conform. Here, I address exclusively behavioral morality, basing my analysis on evolutionary biology, the rational actor model, and behavioral game theory. I suggest that human behavioral morality is the product of an evolutionary dynamic, extending over hundreds of thousands of years in the hominin line, that may be described as culture-led gene–culture coevolution. In this dynamic, hominin societies transformed culture, and the new culture rendered new behaviors fitness enhancing, thereby transforming the hominin line itself. Behavioral morality is thus predicated upon a set of human genetic predispositions that evolved during our evolutionary emergence in small-scale hunter-gatherer groups. When our ancestors developed the intentional capacity to devise social games and play according to these games' culturally constituted rules, it became possible to conceive of society itself as a social game, the rules of which are determined in an emergent arena of social life, which we may call the public sphere. Humans thus evolved two modes of social behavior: a private persona to regulate everyday life in civil society and a public persona to regulate behavior in the public sphere. At the heart of human moral capacities is the capacity to conceptualize a higher moral realm that leads us to behave morally,

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

to feel the satisfaction of behaving morally, and to experience a degraded self when we have not behaved morally.

Gene–Culture Coevolution

Individual fitness in humans depends on the structure of social life. For instance, if social norms entail sanctioning individuals who exhibit certain behaviors, then genes that support these behaviors are likely to be replaced in the population by genes that suppress the sanctioned behaviors.

Human cognitive, affective, and moral capacities are thus the product of an evolutionary dynamic involving the interaction of genes and culture. This dynamic is called gene–culture coevolution (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1982; Dunbar 1993; Richerson and Boyd 2005). This coevolutionary process has endowed us with preferences that go beyond the self-regarding concerns emphasized in traditional economic and biological theory, with a social epistemology that facilitates the sharing of intentionality across minds, and a moral sense that entails both contributing to the social good and doing the right thing for its own sake. Gene–culture coevolution accounts for the salience of such other-regarding values as a taste for cooperation, fairness and retribution, the capacity to empathize, and the ability to value such character virtues as honesty, hard work, piety, and loyalty.

There are strong interactions between genes and culture, ranging from basic physiology, such as the transformation of the organs of speech with the evolution of language, to sophisticated social emotions, including empathy, shame, guilt, and revenge-seeking (Ihara 2011; Zajonc 1980). Because of their common informational and evolutionary character, strong parallels exist between models of genetic and cultural evolution (Mesoudi et al. 2006). Like genes, culture is transmitted from parents to offspring; like culture, which is transmitted horizontally to unrelated individuals as are genes in microbes and many plant species, genes are regularly transferred across lineage boundaries (Jablonka and Lamb 1995; Abbott et al. 2003; Rivera and Lake 2004).

Culture to Genes: The Physiology of Communication

The evolution of the physiology of speech and facial communication is a dramatic example of gene–culture coevolution. The increased social importance of communication in human society rewarded genetic changes that facilitate speech. Regions in the motor cortex expanded in early humans to facilitate speech production. Concurrently, nerves and muscles to the mouth, larynx, and tongue became more numerous to handle the complexities of speech (Jurmain et al. 1997). Parts of the cerebral cortex, Broca's and Wernicke's areas, which do not exist or are relatively small in other primates, are large in humans and

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

permit grammatical speech and comprehension (Belin et al. 2000; Binder et al. 1997).

Modern humans have a larynx low in the throat, a position that allows the throat to serve as a resonating chamber capable of a great number of sounds (Relethford 2007). The first hominids that have skeletal structures supporting this laryngeal placement are the *Homo heidelbergensis*, who lived from 800,000 to 100,000 years ago. In addition, the production of consonants requires a short oral cavity; our nearest primate relatives have much too long an oral cavity for this purpose. The position of the hyoid bone, which is a point of attachment for a tongue muscle, developed in *H. sapiens* in a manner permitting highly precise and flexible tongue movements.

Another indication that the tongue has evolved in hominids to facilitate speech is the size of the hypoglossal canal, an aperture that permits the hypoglossal nerve to reach the tongue muscles. This aperture is much larger in Neanderthals and humans than in early hominids and nonhuman primates (Dunbar 2005). Human facial nerves and musculature also evolved to facilitate communication. This musculature is present in all vertebrates; in mammals, however, it solely serves feeding and respiratory functions (Burrows 2008). In mammals, this mimetic musculature attaches to the skin of the face, thus permitting the facial communication of such emotions as fear, surprise, disgust, and anger. In most mammals, however, a few wide sheet-like muscles are involved, rendering fine information differentiation impossible, whereas in primates, this musculature divides into many independent muscles with distinct points of attachment to the epidermis, thus permitting higher bandwidth facial communication. Humans have the most highly developed facial musculature by far of any primate species, with a degree of involvement of lips and eyes that is not present in any other species.

In short, humans have evolved a highly specialized and very costly array of physiological characteristics that both presuppose and facilitate sophisticated vocal and visual communication, whereas communication in other primates, lacking as they are in cumulative culture, goes little beyond simple calling and gesturing capacities involving adoption of communicative physiology. This example is quite a dramatic and concrete illustration of the intimate interaction of genes and culture in the evolution of our species.

The Rationality of Morality

Behavioral morality involves making personally costly choices that promote ethical goals. People not only balance self-regarding against moral concerns, they also face conflicting moral principles in making choices. Choice behavior is thus modeled using the rational actor model, according to which individuals have a preference function representing their goals, they face constraints that limit the mix of goals available to them, and they have beliefs concerning

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

how their actions affect the probability of attaining their goals. This concept of rationality embodies the consistency principles of formal rationality with little regard for the actor's substantive rationality; that is, the extent to which behavior is efficient in attaining some particular goal state, such as biological fitness, personal well-being, or subjective happiness. Preferences may include self-regarding goals (e.g., material wealth and leisure), other-regarding goals (e.g., fairness, consideration for the welfare of others), and character virtues (e.g., honesty, loyalty, trustworthiness, courage, and considerateness) that have intrinsic value independent of their effects. Moreover, we impose no plausibility constraints on beliefs.

The most important contribution to the theory of formal rational choice was that of Leonard Savage (1954), who showed that a small set of plausible choice axioms (the Savage axioms) implies that a rational actor can be modeled as though he were maximizing a preference function subject to the constraints he faces, where his beliefs take the form of a subjective prior specifying the agent's judgment as to the probabilistic effects of his actions on the attainment of his goals. Modeling the decision maker as a preference function maximizer is a great analytical convenience, but does imply that some conscious or intentional maximization motive exists in the decision maker. The preference function is often called a utility function, although the term is misleading because the preference function in the rational actor model need not have any utilitarian content. The most important of the Savage axioms is that the agent's preferences be transitive in the sense that if he prefers A to B and he also prefers B to C, then he must also prefer A to C. What is extremely important is the no wishful-thinking assumption which states, roughly speaking, that when a choice entails a probability distribution over several possible outcomes, an agent's degree of like or dislike of any one of these outcomes does not affect his assessment of the probability that this outcome will occur. (For a discussion of this principle, which is often violated in highly charged situations, see Gintis 2009a:15; Gintis and Helbing 2014). Often wishful thinking takes the form of maintaining a cherished belief in the face of evidence contradictory to this belief by injudiciously rejecting the new evidence. The remaining assumptions are rather technical and not relevant for our purposes.

The Savage axioms do not suggest that an agent chooses what is in his best interest or what gives him pleasure. Nor do the axioms suggest that the actor is selfish, calculating, or amoral. Finally, the Savage axioms do not suggest that the rational actor is trying to maximize utility or anything else. The maximization formulation of rational choice behavior is simply an analytical convenience. The theory flowing from the Savage axioms is a powerful tool that is valid whatever the nature of human goals and motivations, provided they involve consistent choices.

The rational actor model is most powerful when applied to the analysis of habitual decisions, such as what to purchase at the supermarket, what route to take to work, and which friends to invite to dinner. When a decision is

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

nonhabitual, involving novel elements outside the agent's personal experience, the notion that an agent must deploy his subjective prior in assessing appropriate behavior is inaccurate. In fact, humans are an extremely social species, and each individual is normally embedded in a complex social network including family, friends, coworkers, neighbors, and social media. Individuals thus have networked minds with cognition distributed across the network (Dunbar et al. 2010). When an unfamiliar scenario presents itself, the decision maker will draw on the past experience and embedded beliefs of this network of minds in making a decision. This notion is an extension of the theory of case-based decision making proposed by Gilboa and Schmeidler (2001).

A Typology of Rational Behavioral Morality

Human actors exhibit three types of motives in their daily lives: self-regarding, other-regarding, and universalist. Self-regarding motives include seeking wealth, consumption, leisure, social reputation, status, esteem, and other markers of personal advantage. Other-regarding motives involve a concern for fairness and a compassionate interest in the well-being of others. Universalist motives are those moral rules that are followed for their own sake rather than directly for their effects. Among these universalist goals, which are also termed character virtues, are honesty, loyalty, courage, trustworthiness, and considerateness. Of course, such universalist goals normally have consequences for those with whom one interacts, and for society as a whole. One undertakes, though, universalist actions for their own sake, beyond any consideration of their effects. I will give one example of other-regarding behavior and another of universalist behavior, as revealed by laboratory experiments using behavioral game theory.

Positive Reciprocity: The Trust Game

Positive reciprocity takes the form of an individual responding to an act of kindness by returning the kindness. Positive reciprocity can be self-regarding because returning favors helps create and sustain a mutually rewarding relationship. Trivers (1971) called such tit-for-tat behavior reciprocal altruism, but there is in fact no altruism at all involved, since a purely selfish individual will engage in this form of positive reciprocity. However, humans also exhibit positive reciprocity when there is no possibility of future gain from the costly act of returning a kindness. This other-regarding behavior is called altruistic cooperation.

Consider, for example, the trust game, first studied by Berg et al. (1995). In this game, carried out in an experimental laboratory, subjects are each given an endowment, say $10. Subjects are then randomly paired, and one subject in each pair (the Proposer) is told that he can transfer any number of dollars,

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

from zero to ten, to his anonymous partner (the Respondent) and the Proposer can keep the remainder. The amount transferred will be tripled by the experimenter and given to the Respondent, who can then give any number of dollars back to the Proposer (this amount is not tripled). A Proposer who transfers a lot is called trusting; a Respondent who returns a lot to the Proposer is called trustworthy. This interaction occurs only one time, and the Proposer and the Respondent never learn each other's identity. Trustworthiness is thus a pure act of other-regarding positive reciprocity.

On average, Berg et al. (1995) found that the Proposer transferred $5.16 of the $10.00 to the Respondent, and on average, the Respondent transferred $4.66 back to the Proposer. Furthermore, when the experimenters revealed this result to the subjects and had them play the game a second time, on average $5.36 was transferred from the Proposer to the Respondent, and $6.46 was transferred back from the Respondent to the Proposer. In both sets of games there was a great deal of variability: some Proposers transferred everything while some gave nothing, and some Respondents more than fully repaid their Proposers while others returned nothing.

Negative Reciprocity: The Ultimatum Game

Negative reciprocity occurs when an individual responds to an unkind act by retaliating with another unkind act. Negative reciprocity can be self-regarding because retaliation may induce the other person to behave more kindly in the future, thereby enhancing one's reputation as someone not to be trifled with. There is no moral element in this sort of negative reciprocity, since a purely selfish individual may retaliate to enhance his reputation and thereby deter future unkind acts. However, humans also exhibit negative reciprocity when there is no possibility of future interaction with the offender. This other-regarding negative reciprocity is called altruistic punishment.

The simplest game exhibiting altruistic punishment is the Ultimatum Game (Güth et al. 1982). Under conditions of anonymity, two subjects, whom we will call Alice and Bob, are shown a sum of money, say $10.00. Alice (the Proposer) is instructed to offer any number of dollars, from $1.00 to $10.00, to Bob (the Responder). Alice can make only one offer, and Bob can either accept or reject this offer. If Bob accepts the offer, the money is split according to Alice's offer. If Bob rejects the offer, both players receive nothing. Alice and Bob, who are unknown to each other, do not interact again.

If Bob is self-regarding, he will accept anything offered. If Alice believes Bob is self-regarding, she will offer him the minimum amount ($1.00) and Bob will accept. However, when actually played, this self-regarding outcome is almost never observed or even approximated. In fact, under varying conditions and with varying amounts of money, Proposers routinely offer Responders very substantial amounts (50% of the total generally being the modal offer)

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

and Responders frequently reject offers below 30% (Güth and Tietz 1990; Camerer and Thaler 1995).

Are these results culturally dependent? Do they have a strong genetic component or do all successful cultures transmit similar values of reciprocity to individuals? Roth et al. (1991) conducted the Ultimatum Game in four different countries (the United States, the former Yugoslavia, Japan, and Israel) and found that while the level of offers differed a small but significant amount in different countries, the probability of an offer being rejected did not. This indicates that both Proposers and Responders share the same notion of what is considered fair in that society, and that Proposers adjust their offers to reflect this common notion. When a much greater degree of cultural diversity is studied, however, large differences in behavior are found, reflecting different standards of what it means to be fair in different types of societies (Henrich et al. 2004).

Behavior in the Ultimatum Game conforms to the altruistic punishment model. Responders reject offers under 40% to hurt an unfair Proposer. Proposers offer 50% because they are altruistic cooperators, or 40% because they fear rejection.

To support this interpretation, if the offers in an Ultimatum Game are generated by a computer rather than by the Proposer, and if Responders know this, low offers are rarely rejected (Blount 1995). This suggests that players are motivated by reciprocity, reacting to a violation of behavioral norms (Greenberg and Frisch 1972). Moreover, in a variant of the game in which a Responder rejection leads to the Responder getting nothing but allows the Proposer to keep the share he suggested for himself, Responders never reject offers, and proposers make considerably smaller (but still positive) offers (Bolton and Zwick 1995). As a final indication that altruistic punishment motives are operative in this game, after the game is over, when asked why they offered more than the lowest possible amount, Proposers commonly said that they were afraid that Responders will consider low offers unfair and reject them. When Responders rejected offers, they usually claimed they wanted to punish unfair behavior. In all of the above experiments, a significant fraction of subjects (about a quarter, typically) conformed to purely self-regarding preferences.

A Universalist Character Virtue: Honesty

Certain moral behaviors are universalist in the sense that one performs them, at least in part, because it is virtuous to do so, apart from any effects they have on oneself, others, or society in general. For instance, one can be honest in dealing with another agent without caring at all about the effect on the other agent, or even caring about the impact of honest behavior on society at large. Similarly, one can be courageous in battle because it is the right thing to do, independent of the effect of one's actions on winning or losing the battle.

A particularly clear example of the value of honesty was reported by Gneezy (2005), who studied 450 undergraduate participants paired off to play three

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

games where all payoffs took the form (a; b): player 1 (Alice) receives a and player 2 (Bob) receives b. In all games, Alice was shown two pairs of payoffs, A:(x; y) and B:(z; w) where x, y, z, and w are amounts of money with $x < z$ and $y > w$, so in all cases, B is better for Bob and A is better for Alice. Alice could then say to Bob, who is unable to see the amounts of money: "Option A will earn you more money than option B," or "Option B will earn you more money than option A." The first game was A:(5; 6) versus B:(6; 5), so Alice could gain one by lying and being believed, while imposing a cost of one on Bob. The second game was A:(5; 15) versus B:(6; 5), so Alice could gain ten by lying and being believed, while still imposing a cost of one on Bob. The third game was A:(5; 15) versus B:(15; 5); here Alice could gain ten by lying and being believed, while imposing a cost of ten on Bob.

Before starting to play, the experimenter asked Alice whether she expected her advice to be followed, inducing honest responses by promising to reward her if her guesses were correct. The experimenter found that 82% of Alices expected their advice to be followed (the actual result was that 78% of Bobs followed their Alice's advice). It follows that if Alices were self-regarding, they would always lie and recommend B to their Bob.

The experimenters found that in game two, where lying was very costly to Bob and the gain to lying for Alice was small, only 17% of subjects lied. In game one, where the cost of lying to Bob was only one but the gain to Alice was the same as in game two, 36% lied. In other words, subjects were loathe to lie, but considerably more so when it was costly to their partner. In game three, where the gain from lying was large for Alice, and equal to the loss to Bob, fully 52% lied. This shows that many subjects are willing to sacrifice material gain to avoid lying in a one-shot, anonymous interaction, their willingness to lie increasing with an increased cost of truth-telling to themselves, and decreasing with an increase in their partner's cost of being deceived. Similar results were obtained by Boles et al. (2000) and Charness and Dufwenberg (2006). Gunnthorsdottir et al. (2002) and Burks et al. (2003) have shown that a social-psychological measure of "Machiavellianism" predicts which subjects are likely to be trustworthy and trusting.

The Public Sphere

The social life of most species, including mating practices, symbolic communication, and power relations, is inscribed in its core genome and expressed in stereotypical form by its members (Gintis 2014). *H. sapiens* is unique in adapting its social life in fundamental and deep-rooted ways to environmental challenges and opportunities (Richerson and Boyd 2005). This flexibility is based on two aspects of our mental powers. The first is our ability to devise new rules of the game in social life, and to base our social interaction on these new rules. This capacity, absent in other species, makes us *H. ludens*: man the game

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

player. This capacity is possessed even by very young children who invent, understand, and play games for fun. In adult life, this same capacity is exercised when people come together to erect, protect, and transform the social rules that govern their daily lives. Broadly speaking, we can define the public sphere as the arena in which society-wide rules of the game are considered, and politics as the cooperative, conflictual, and competitive behaviors through which rules are established and individuals are assigned to particular public positions.

Humans evolved in hunter-gatherer societies consisting of a dozen families or so (Kelly 1995), in which political life was an intimate part of daily life, involving the sorts of self-regarding, other-regarding, and universalistic motivations described above. These individual bands were embedded in a larger ethnolinguistic group consisting of many hundreds or even thousands of individuals. In such a setting, political activity was strongly consequentialist: a single individual could expect to make a difference to the outcome of a deliberation, a conflict, or a collaboration, so that our political morality developed intimately entwined with material interests and everyday consequentialist moral sentiments (Boehm 1999, 2012; Gintis et al. 2015).

As we move from small-scale hunter-gatherer societies to modern mass societies with millions of members, the public sphere passes from being intimately embedded in daily life to being a largely detached institutional arena, governed by complex institutions controlled by a small set of individuals, and over which most members have at best formal influence through the ballot box, and at worst no formal influence whatever. Political activity in modern societies is thus predominately nonconsequentialist, meaning that individuals do not base their choices on the effect of their actions on political outcomes (Quattrone and Tversky 1998; Shayo and Harel 2012). Except for a small minority of individuals contesting for personal power, the political choices of a single citizen affects public sphere outcomes with a probability very close to zero—sufficiently close that these choices cannot be attributed to consequentialist motives, whether self-regarding, other-regarding, or universalist.

In large elections, the rational consequentialist agent will not vote because the costs of voting are positive and significant, but because the probability that one vote will alter the outcome of the election is vanishingly small, and adding a single vote to the total of a winning candidate enhances the winner's political efficacy at best an infinitesimal amount (Downs 1957; Riker and Ordeshook 1968). Thus the personal consequentialist gain from voting is too small to motivate behavior even for a committed other-regarding or universalist altruist (Hamlin and Jennings 2011). For similar reasons, if one chooses to vote, there is no plausible reason to vote on the basis of the impact of the outcome of the election on one's personal material gains, or on the basis of the gains to the demographic and social groups to which one belongs, or even on the basis of consequentialist universal values. One vote simply makes no difference. It follows also that the voter, if rational and consequentialist, and incapable of personally influencing the opinions of more than a few others, will not bother

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

to form opinions on political issues, because these opinions cannot affect the outcome of elections. Yet people do vote, and many do expend time and energy in forming political opinions. Although voters do appear to behave strategically (Fedderson and Sandroni 2006), their behavior does not conform to the rational consequentialist model (Edlin et al. 2007).

It also follows that rational consequentialist individuals will not participate in the sort of collective actions that are responsible for the growth in the world of representative and democratic governance, the respect for civil liberties, the rights of minorities and gender equality in public life, and the like. In the rational consequentialist model, only small groups aspiring for social dominance will act politically. Yet modern egalitarian political institutions are the result of such collective actions (Bowles and Gintis 1986; Giugni et al. 1998). This behavior cannot be explained by a rational consequentialist model.

Politically active and informed citizens appear to operate on the principle that voting and participating in collective actions are highly valued nonconsequentialist behaviors. This idea is difficult for people to articulate because the consequentialist versus nonconsequentialist distinction is not part of either common parlance or the specialized lexicon of political theory. However, most voters agree with statements like "my single vote won't make a difference, but if all concerned citizens vote our common concerns, we can make a difference." Of course it does not logically follow that one should vote according to standard decision theory because if "my single vote won't make a difference," then I still have no consequentialist reason for voting.

Humans, however, appear to follow a nonconsequentialist logic that may be described as distributed effectivity: act in a way that best reflects your preferences assuming that your choices are consequential even when they are not. For instance, decide whether or not to vote, and how to cast your ballot, assuming that the electorate is so small that your decision may be pivotal in deciding the outcome of the election (Gintis 2015). For instance, suppose agent i has payoff b_i if his side wins the election and zero if it loses, and i's cost of voting is c_i, then i will vote provided that

$$b_i p > c_i, \tag{7.1}$$

where p is the probability of being a pivotal voter, thus swinging the election from loss to win. In a large election, p will be infinitesimal and Equation (7.1) will not be satisfied. But suppose distributed effectivity reasoning dictates that we act prosocially, as though the electorate were tiny (say fifty potential voters), then Equation (7.1) can be satisfied with quite large voting cost (Gintis 2015). Note that b_i in this equation need not reflect selfish behavior, but rather other-regarding and universalist as well. Moreover c_i need not be restricted to material costs, but could include feelings of duty and social responsibility or a wish to signal one's social value.

Distributed effectivity is related to rule consequentialism (Harsanyi 1977; Hooker 2011; Roemer 2010), and team-based reasoning (Bacharach 2006;

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

Gilbert 1989; Searle 1995; Tuomela 1995). Each of these terms captures part of the somewhat subtle, yet substantively important, notion that human minds often act together even when classical rational choice theory says they will not. Rule consequentialism is the principle that like-minded agents become consequential by virtue of the common decision algorithms and values. Team reasoning carries the connotation that the agent is both privileged to be part of a social process, so that the collective goal becomes internalized as a personal goal, and yet is duty-bound and committed to performing his part in the social decision process even when this conflicts with personal goals. By contrast, distributed effectivity is a purely personal decision process with a distinct moral dimension: reason as though your choices matter.

Indeed, the voting public regularly recognizes that individual votes do not matter, yet voting is an important social activity. In one Wisconsin survey reported by Dennis (1970), 89% of respondents disagreed with the statement, "So many other people vote in the national elections that it doesn't matter much to me whether I vote or not." Moreover, 86% of respondents agree with the statement, "I generally get a feeling of satisfaction from going to the polls to cast my ballot," and 65% of respondents disagree with the statement, "A person should only vote in an election if he cares about how it is going to come out."

The validity of distributed effectivity implies that people are perfectly reasonable in assenting to such assertions as "I am helping my candidate win by voting" and "I am helping promote democracy by demonstrating against the dictator." These assertions are rigorously correct, despite the nonconsequential nature of the acts of individual minds. Because distributed effectivity is so ingrained in our public persona, people untrained in traditional rational decision theory simply cannot understand the argument that it is irrational to vote or to participate in collective actions, even when they can be persuaded that their actions are in fact nonconsequential.

It might be entertained that the preference for political activity is simply a form of human expression without close ties to decision theory (Hamlin and Jennings 2011). However, distributed effectivity is compatible with many stylized facts of voter behavior (Gintis 2015; Levine and Palfrey 2007) that are clearly related to rational decision making. In general, we can describe the dynamics of distributed effectivity by modeling even very large elections as though they were very small elections with a few dozen participants. The first fact is the voting cost effect: when the cost of voting increases, fewer people vote. The reason is that each voter has a maximum cost of voting threshold that depends on the strength of his commitment to the political process, his income and wealth, and the details of his immediate obligations and opportunities. These costs of voting thresholds will have a statistical distribution such that a general rise in the cost of voting (e.g., bad weather) will lead to lower voter turnout. The second is the outcome importance effect: voter turnout is higher when the issues to be decided have greater social impact. The reason is that in a more important election, the ratio of cost of voting to benefits of winning the

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

election will fall, thus increasing the number of agents who will vote. The third is the close election effect: turnout is higher when the election is expected to be close. The reason is that in a close election a small number of voters is likely to tip the outcome one way or the other, so agents feel an increased commitment to vote (Shachar and Nalebuff 1999). The fourth is the underdog effect: in a two-party election, turnout is generally higher among voters for the side that is not expected to win. This is because a minority voter is more likely to be a pivotal voter than a majority voter. Of course, in a very large election, both probabilities would be so small as to be negligible. But with distributed effectivity, agents reason that their choice is diagnostic of what other voters will do, so the relevant probabilities become consequential. The fifth is that strategic voting is widely observed, such as ignoring candidates that have no hope of winning, and voting for an unwanted candidate in order to avoid electing an even less preferred candidate (Cox 1994; Niemi et al. 1992). Finally, in many circumstances there is a strong social network effect: individuals who are more solidly embedded in strong social networks tend to vote at a higher rate (Edlin et al. 2007; Evren 2012). This is because the power of distributed effectivity is a function of the complexity of the network of distributed cognition to which the individual belongs.

We conclude that the individual immersed in consequentialist everyday life expresses his private persona, while his behavior in the public sphere reveals his public persona. Individuals acting in the public sphere are, then, a different sort of animal, one which Aristotle called *zoon politikon* in his Nicomachean Ethics in 350 BC.

Private and Public Persona

The concept of a nonconsequentialist public persona suggests a two by three categorization of human motivations, as presented in Table 7.1. In this table, the three columns represent three modes of social interaction. The personal mode is purely self-regarding, while the social mode represents the agent as embedded in a network of significant social relations, and the universal represents the individual's realm of recognized supra-situational moral obligations. The two rows represent the agent's private persona of consequentialist social relations in civil society, and the agent's public persona of nonconsequentialist political relationships in the public sphere.

Table 7.1 A typology of human motivations.

	Personal	Social	Universal
Private persona	*Homo economicus*	*Homo socialis*	*Homo vertus*
Public persona		*Homo parochialis*	*Homo universalis*

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

H. economicus is the venerable rational selfish maximizer of traditional economic theory; *H. socialis* is the other-regarding agent who cares about fairness, reciprocity, and the well-being of others; and *H. moralis* is the Aristotelian bearer of nonconsequentialist character virtues. The new types of public persona are *H. parochialis*, who votes and engages in collective action on behalf of the narrow interests of the demographic, ethnic and/or social status groups with which he identifies (Ben-Bassat and Dahan 2012; Coate and Conlin 2004), and *H. universalis*, who acts politically to achieve what he considers the best state for the larger society, for instance, reflecting John Rawls's (1971) veil of ignorance, John Harsanyi's (1977) criterion of universality, or John Roemer's (2010) Kantian equilibrium.

The public persona/personal box is empty because a self-regarding agent will never do anything except for its consequences, and the concept of distributed effectivity will not apply. Interestingly, the individual whose private persona is social is generally considered altruistic, whereas the individual whose public persona is social is often considered selfish, acting in a partisan manner on behalf of the narrow interests of the social networks to which he belongs. Of course *H. parochialis* is in fact altruistic toward these social networks.

The Evolutionary Emergence of Private Morality

By cooperation we mean engaging with others in a mutually beneficial activity. Cooperative behavior may confer net benefits on the individual cooperator, and thus can be motivated entirely by self-interest. In this case, cooperation is a form of mutualism. Cooperation may also be a net cost to the individual but the benefits may accrue to a close relative. We call this kin altruism. Cooperation can additionally take the form of one individual's costly contribution to the welfare of another individual being reliably reciprocated at a future date. This is often called reciprocal altruism (Trivers 1971), although it is non-altruistic tit-for-tat mutualism. However, some forms of cooperation impose net costs upon individuals, the beneficiaries may not be close kin, and the benefit to others may not be expected to be repaid in the future. This cooperative behavior is true altruism.

The evolution of mutualistic cooperation and kin altruism are easily explained. Cooperation among close family members evolves by natural selection because the benefits of cooperative actions are conferred on the close genetic relatives of the cooperator, thereby helping to proliferate genes associated with the cooperative behavior. Kin altruism and mutualism explain many forms of human cooperation, particularly those occurring in families or in frequently repeated two-person interactions. But these models fail to explain two facts about human cooperation: that it takes place in groups far larger than the immediate family, and that both in real life and in laboratory experiments, it occurs in interactions that are unlikely to be repeated, and where it is impossible

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

to obtain reputational gains from cooperating. These forms of behavior are regulated by moral sentiments.

The most parsimonious proximal explanation of altruistic cooperation, one that is supported by extensive experimental and everyday-life evidence, is that people gain pleasure from cooperating and feel morally obligated to cooperate with like-minded people. People also enjoy punishing those who exploit the cooperation of others. Free riders frequently feel guilty, and if they are sanctioned by others, they often feel ashamed. We term these feelings social preferences. Social preferences include a concern, positive or negative, for the well-being of others, as well as a desire to uphold social norms.

The Roots of Social Preferences

Why are the social preferences that sustain altruistic cooperation in daily life so common? Early human environments are doubtless part of the answer. Our Late Pleistocene ancestors inhabited the large-mammal-rich African savannah and other environments in which cooperation in acquiring and sharing food yielded substantial benefits at relatively low cost. As a result, members of groups that sustained cooperative strategies for provisioning, child-rearing, sanctioning noncooperators, defending against hostile neighbors, and truthfully sharing information had significant advantages over members of noncooperative groups.

There are several reasons why these altruistic social preferences supporting cooperation outcompeted amoral self-interest. First, human groups devised ways to protect their altruistic members from exploitation by the self-regarding. Prominent among these is the collective punishment of miscreants (Boyd et al. 2010), including the public-spirited shunning, ostracism, and even execution of free riders and others who violate cooperative norms.

Second, humans adopted elaborate systems of socialization that led individuals to internalize the norms that induce cooperation, so that contributing to common projects and punishing defectors became objectives in their own right rather than constraints on behavior. Together, the internalization of norms and the protection of the altruists from exploitation served to offset, at least partially, the competitive handicaps born by those who were motivated to bear personal costs to benefit others.

Third, between-group competition for resources and survival was and remains a decisive force in human evolutionary dynamics. Groups with many cooperative members tended to survive these challenges and to encroach upon the territory of the less cooperative groups, thereby both gaining reproductive advantages and proliferating cooperative behaviors through cultural transmission. The extraordinarily high stakes of intergroup competition and the contribution of altruistic cooperators to success in these contests meant that sacrifice on behalf of others, extending beyond the immediate family and even to virtual strangers, could proliferate (Bowles 2009; Choi and Bowles 2007).

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

This is part of the reason why humans became extraordinarily group-minded, favoring cooperation with insiders and often expressing hostility toward outsiders. Boundary maintenance supported within-group cooperation and exchange by limiting group size and within-group linguistic, normative, and other forms of heterogeneity. Insider favoritism also sustained the between-group conflicts and differences in behavior that made group competition a powerful evolutionary force.

The Evolutionary Emergence of the Public Persona

Nonhuman species, even if highly social, do not engage in intentional activities that structure the social rules that regulate their lives. Therefore there is no politics and no public sphere in these species, and hence its members have no public persona. How, then, might a public persona have arisen in the hominin line leading up to *H. sapiens*?

My colleagues and I supply an answer grounded in the information available to us from a variety of fields, including paleontology, primatology, the anthropology of contemporary hunter-gatherer groups, animal behavior theory, and genetics (Gintis et al. 2015). We propose that the emergence of bipedalism, cooperative breeding, and lethal weapons (stones and spears) in the hominin line, together with favorable climate change, made the collaborative hunting and scavenging of large game fitness enhancing. Lethal weapons are the most unique of these innovations, as other predators (e.g., lions, tigers and other big cats, wolves, foxes, and other canines) use only their natural weapons (sharp claws and teeth, powerful jaws, great speed) in hunting, while none of these endowments was available to early hominins.

Lethal hunting weapons, moreover, transformed human sociopolitical life because they could be applied to humans just as easily as to other animals. Indeed, the combination of the need for collaboration and the availability of lethal weapons in early hominin society undermined the social dominance hierarchy characteristic of primate and earlier hominin groups, which was based on pure physical prowess. The successful sociopolitical structure that ultimately replaced the ancestral social dominance hierarchy was an egalitarian political system in which lethal weapons made possible group control of leaders, group success depended on the ability of leaders to persuade and motivate, and of followers to contribute to a consensual decision process. The heightened social value of nonauthoritarian leadership entailed enhanced biological fitness for such leadership traits as linguistic facility, ability to form and influence coalitions, and indeed for hypercognition in general.

This egalitarian political system persisted until cultural changes in the Holocene fostered the accumulation of material wealth, through which it became possible once again to sustain a social dominance hierarchy with strong authoritarian leaders based on physical coercion.

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.

Conclusion

This chapter has provided evidence for a model of human behavior based on the rational actor model, in which individuals have both private and public persona, and their preferences range over personal, social, and universal modes of private persona and in most activities in the public sphere. Morality in this model is defined in behavioral terms: moral choices are those made in social and universalist modes. The public sphere in this model is an arena where preferences and actions are primarily nonconsequentialist. The other-regarding preferences of *H. socialis* and the character virtues of *H. vertus* are underpinnings of civil society, whereas *H. parochialis* and *H. universalis* make possible the varieties of political life characteristic of our species in the modern era.

From "Complexity and Evolution: Toward a New Synthesis for Economics,"
David S. Wilson and Alan Kirman, eds. 2016. Strüngmann Forum Reports, vol. 19,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03538-5.