

9 Morality: Evolutionary Foundations and Policy Implications

Ingela Alger and Jörgen W. Weibull

Act only according to that maxim whereby you can, at the same time, will that it should become a universal law.

—Immanuel Kant, *Groundwork of the Metaphysics of Morals*, 1785

One general law, leading to the advancement of all organic beings, namely, multiply, vary, let the strongest live and the weakest die.

—Charles Darwin, *On the Origin of Species*, 1859

The academic discipline of economics has over many years provided policy makers all over the world with a powerful toolbox. Conceptual, philosophical, and methodological disagreements are relatively rare, and the discipline is not torn by fights between disparate schools of thought. Whether this monolithic character of the field is a sign of strength or weakness is not easy to say, but this methodological unity and power has, arguably, given the discipline great influence on policy. The strong methodological core of economics—in the 1950s–1960s epitomized by general equilibrium theory and later incorporating game theory—has enabled positive and normative analysis of a wide range of economic and social issues.

So what, more exactly, does this core consist of? In a nutshell, it has two main components. The first is that it views economic agents—who may be individuals, households, firms, or organizations—as goal-oriented,

This manuscript was prepared for the conference “The State of Economics, The State of the World,” held at the World Bank in Washington, DC, on June 8–9, 2016. The authors thank Daniel Chen, Jean-François Laslier, Assar Lindbeck, Erik Mohlin, Paul Seabright, Jean Tirole, Nicolas Treich, Yu Wen, and Peter Wikman for valuable comments and suggestions.

as if they each had some goal function that they strive to maximize under the constraints they face, given the information they have, and given their beliefs about relevant aspects of the world they live in. The second component is that interactions between these economic agents are taken to meet certain consistency requirements, formalized as equilibria, that is, collections of action plans, one for each agent, such that no agent can unilaterally improve the expected value of her goal function (usually profit or utility).

Both components can and have been contested. Individuals may not be so systematic and consistent, and interactions may be chaotic and volatile. Having a theoretically well-founded and empirically accurate understanding of human motivation is, arguably, in any case of utmost relevance for analysis and policy recommendations.

Among the more noticeable new methodological developments in economics is the emergence of behavioral and experimental economics, where the first strand endows economic agents with richer motivations than in traditional economics, usually in the form of prosocial or other-regarding preferences. The second strand tests such models, old and new, in controlled laboratory experiments and in randomized field experiments. The external validity of laboratory experiments can be questioned, and field experiments may depend on local and historical factors with little generality, but this development of the discipline of economics toward becoming an empirically founded science appears to be essentially very healthy. It was not long ago that economics was thought of as similar to meteorology and astronomy: All it could do was to observe what is happening, without the possibility of experimenting. Moving away from mere observation of data that happen to come about to carefully designed controlled experiments is reminiscent of how Galileo Galilei once lead the way from Aristotelean scholastic discourse to modern science.

Behavioral and experimental economics no doubt will improve the predictive power and the usefulness of economics, but further improvements could certainly be made if the underlying factors that shape human motivation were better understood. The literature on the evolutionary foundations of human motivation aims at providing such understanding by asking: What preferences should humans be expected to have if these are transmitted in society from generation to generation? If certain prosocial or antisocial preferences, or moral values, give their carriers on average better material outcomes than other preferences or values (all else being equal),

then one would expect the former to spread in the population (be it by biological or cultural mechanisms). Our aim in this chapter is to discuss a recent theoretical result concerning such evolutionary preference selection and to examine its implications for a range of social and economic issues.

Milton Friedman (1953, 22) claimed that “unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long.” In a similar vein, one may claim that unless the behavior of an individual is consistent with the maximization of own material payoffs, other, materially more successful behaviors will take over in the interacting population. Economists have shown that this claim is theoretically valid when (1) the population at hand is very large, (2) interacting individuals do not know each other’s goal functions, and (3) interactions are perfectly random in the sense that each encounter is just as likely (Ok and Vega-Redondo 2001; Dekel, Ely, and Yilankaya 2007).

In reality, however, populations are not always large, and interacting individuals sometimes know or learn about each other’s preferences (for instance, think of the great number of interactions that take place in families or small communities). It has been shown that in such settings, preferences or goal functions can usually serve as effective commitment devices, and evolution will almost always favor goal functions that differ from own material payoffs.¹ Furthermore—and this is what we will focus on here—encounters are only rarely perfectly random; geographic location, language, culture, and religion often have an impact on the likelihood of specific encounters. For example, business partners may know each other from college, and neighbors may have chosen to live in the same place because they share socioeconomic or cultural backgrounds or have similar location preferences and so forth. In such structured populations, some encounters are more likely than others, even if the overall population is large. In two recent theoretical studies (Alger and Weibull 2013, 2016), we show that such *assortative matching* makes evolution favor individuals who are not purely self-interested but who attach some value to “doing the right thing,” even though the population is large, and interacting individuals

1. Seminal articles on preference evolution, or indirect evolution, are Frank (1987) and Güth and Yaari (1992). See also Banerjee and Weibull (1995); Heifetz, Shannon, and Spiegel (2007), and Alger and Weibull (2010).

do not know one another's preferences. This (for us, initially surprising) finding suggests an evolutionary foundation for a psychologically plausible form of morality, in line with Immanuel Kant's categorical imperative.

In the next section, we describe this novel class of preferences and their evolutionary foundations. In the second section, we discuss the implications of such preferences for some much-studied social and economic behavior and policy issues, including public goods provision and behaviors that affect the environment. The third section discusses other social preferences and contrasts morality with altruism. The final section concludes.

Evolution and Kantian Morality

Imagine a population that has evolved for many generations in a stationary environment and that in each generation, individuals engage in some social or economic interaction. For instance, in a population of self-subsistence farmers, the interaction could be teamwork in the fields, the extraction of resources from a commonly owned lake or piece of land, lending activities, or the maintenance of institutions. In Alger and Weibull (2013, 2016), we propose a theoretical model of precisely such populations. We formalize the interaction by assuming that individuals are now and then randomly matched into groups of arbitrary (but fixed and given) size n to interact with each other in the group. (There are no interactions between groups and hence no group selection takes place.) The interaction may involve elements of cooperation and/or conflict, asymmetric information, repetition or interaction of arbitrary duration, possibility of helping, rewarding and/or punishing others, and so forth. Essentially only two restrictions are imposed on the interaction. First, the material payoff consequences for a participant depend only on the participant's own actions and on some aggregate of other group members' actions (not on who of them does what). In game theory, such interactions are called *aggregative* games. Examples are market competition where only competitors' aggregate output or lowest price matters, contributions to public goods where only the sum of others' contributions matters, some environmental externalities, and the like. Second, the material payoff function is the same for all individuals.

We follow standard economic theory by assuming that each individual acts so as to maximize some goal function. Different goal functions may be present in the population where each goal function represents some

preference. Depending on the preference distribution and the process by which interaction groups are formed, individuals may end up in more or less homogeneous groups. For a given material interaction, a given preference distribution, and a given group formation process, the average material payoff consequences for individuals with a particular goal function are well determined in each equilibrium. In our evolutionary stability analysis, we ask: What kind of goal function, if any, would be favored by natural selection? Specifically, we determine which such functions are evolutionarily stable in the sense that, if almost all individuals in the population have such preferences, these individuals would materially outperform individuals with other preferences. Thus, the material payoffs are taken to be the drivers of evolution.

This approach is a generalization of the work of Maynard Smith and Price (1973), from the notion of an evolutionarily stable strategy (ESS), to that of an *evolutionarily stable goal function*.² A major challenge arises with this generalization. In any *population state*—the preference distribution in the population—there may be multiple equilibrium behaviors, and hence several possible material payoff allocations. We define a goal function to be *evolutionarily stable against another goal function* if in every population state where the latter goal function is rare, individuals equipped with the former goal function outperform those with the latter in terms of the resulting material payoffs in all equilibria.³ Conversely, a goal function is *evolutionarily unstable* if there exists another goal function such that, no matter how small its population share, there is some equilibrium in which the latter goal function materially outperforms the former. In both definitions, the test scenario is to let in a small population share of “mutants,” who may be migrants or carriers of spontaneously and randomly arising alternative goal functions, into the population of incumbents or residents. We impose minimal constraints on the nature of potential goal functions. They are not required to take any particular parametric form or even to depend on the material payoffs. Hence, individuals may be selfish, altruistic, spiteful,

2. In our approach, it is thus as if “mother nature” delegates to individuals to choose their actions, and instead equips them with goal functions that will guide their choice of action.

3. By “equilibrium” we mean Bayesian Nash equilibrium under incomplete information.

fairness-minded, inequity averse, environmentalists, moralists, and the like. Our only assumption is that each individual's goal function is continuous in all group members' courses of action.

A second key feature of our approach is that it allows the random matching to be *assortative*. Geographic, cultural, linguistic, and socioeconomic distances impose (literal or metaphoric) transportation costs, which imply that (1) individuals tend to interact more with individuals in their (geographic, cultural, linguistic, or socioeconomic) vicinity,⁴ and (2) cultural or genetic transmission of types (say, behavior patterns, preferences, or moral values) from one generation to the next also has a tendency to take place in the vicinity of where the type originated.⁵ Taken together, these two tendencies imply that individuals who interact with each other are likely to be of the same type. We formalize such potential assortativity in the random matching process in terms of a vector we call the *assortativity profile*. This vector consists of probabilities for the events that none, some, or all individuals in a vanishingly rare mutant's group also are mutants.⁶

Our analysis delivers two main results. First, although we impose virtually no restrictions on permissible utility functions, evolution favors a particular class of utility functions, which we call *Homo moralis*. Individuals with preferences in this class attach some weight to their own material payoff but also to what can be interpreted as a probabilistically generalized version of Kantian morality. In his *Grundlegung zur Metaphysik der Sitten*, Immanuel Kant ([1785] 2002, 37) wrote: "Act only according to that

4. Homophily has been documented by sociologists (e.g., McPherson, Smith-Lovin, and Cook 2001; Ruef, Aldrich, and Carter 2003) and economists (e.g., Currarini, Jackson, and Pin 2009, 2010).

5. In biology, the concept of assortativity is known as *relatedness*, and the propensity to interact with individuals locally is nicely captured in the infinite island model, originally due to Wright (1931). Hamilton (1964) provided a first formalization of what is now known as Hamilton's rule: Evolution will select for behaviors whereby the external effects on others are internalized at a rate provided by the relatedness (see also Dawkins (1976), for a popular account of this idea, as well as Rousset (2004), for a comprehensive treatment). In an article on the evolution of behaviors in interactions among siblings, Bergstrom (1995) was probably the first to bring Hamilton's rule into the economics literature.

6. This concept generalizes Bergstrom's (2003) definition of the *index of assortativity* for pairwise encounters. See also Bergstrom (2012) and Alger and Weibull (2013) for further discussions of assortativity under pairwise matchings.

maxim whereby you can, at the same time, will that it should become a universal law." Similarly, paraphrasing Kant, *Homo moralis* attaches some weight to the goal of acting according to that maxim whereby you can, at the same time, will that it should become a universal law, even if followed only probabilistically by others. More precisely, a *Homo moralis* individual in a group of arbitrary size n maximizes a weighted average of equally many terms, indexed $j=0, \dots, n-1$, where each term is the material payoff that she would obtain if, hypothetically, she could replace the strategies of j other individuals in the group by her strategy. We call the vector of these probability weights the individual's *morality profile*.

The class of *Homo moralis* preferences has two extremes: *Homo oeconomicus*, who considers only her own material payoff,⁷ and *Homo kantiansis*, who considers only the material payoff that she would obtain if all others were to act like she does. In between these two extremes is a whole range of *Homo moralis* preferences with different morality profiles, whereby an individual examines what would happen if some but not all the others were to act like him- or herself. *Homo moralis* partly evaluates her own actions in this probabilistic Kantian sense. In other words, she is to some extent concerned with the morality of her own acting, irrespective of what others do. She asks herself, before taking her action, what action she would prefer if, hypothetically, others would also probabilistically choose the same action in her situation.

Our first main result is that *Homo moralis* with a morality profile identical to the assortativity profile is evolutionarily stable. The intuition behind this result is *not* based on group selection, an old argument (appearing already in Charles Darwin's writings; see also Alexander (1987)) that essentially says that evolution will lead to behaviors that enhance the survival of the group. Quite on the contrary; the intuition is that natural selection will lead to utility functions that *preempt entry* into the population in the sense that the best a potential rare mutant can do, if striving for material payoff, is to mimic the residents.

7. Note that we define *Homo oeconomicus* as individuals who always seek to maximize their own material payoff. Some writers define *Homo oeconomicus*, or "economic man," more generally as an individual who always acts in accordance with some goal function, whether this be pure self-interest or not. All agents in the present study are varieties of *Homo oeconomicus* in this broad sense.

Our second main result is that any preferences that are behaviorally distinct from those of *Homo moralis* with the stable morality profile are evolutionarily unstable. Hence, although we made no parametric or structural assumption about utility functions, it appears that natural selection—as represented by evolutionary stability in our abstract and simplified framework—favors the utility function of *Homo moralis*. In particular, our results imply that *Homo oeconomicus*—pure material self-interest—is evolutionarily unstable under any random matching process with positive assortativity. Rare mutants may indeed garner a higher material payoff than *Homo oeconomicus*, on average, by behaving somewhat prosocially, because when there is positive assortativity, the benefits of this prosocial behavior are sometimes bestowed on other mutants, whereas the residents almost never benefit from it.

Homo moralis is easily defined for pairwise interactions, $n=2$. Let $\pi(x, y)$ denote the material payoff to an individual who plays strategy x when the opponent plays strategy y . Then the utility function of *Homo moralis* is

$$U_{\kappa}(x, y) = (1 - \kappa) \cdot \pi(x, y) + \kappa \cdot \pi(x, x), \quad (1)$$

where $0 \leq \kappa \leq 1$ is the individual's *degree of morality*. The two extreme degrees of morality represent *Homo oeconomicus* ($\kappa=0$) and *Homo kantiansis* ($\kappa=1$), respectively, and intermediate degrees of morality correspond to individuals who attach some weight to their own material payoff, $\pi(x, y)$, and some weight to “the right thing to do if everyone were to choose the same behavior,” $\pi(x, x)$.

For $n > 2$, the precise definition of *Homo moralis* is fairly involved,⁸ but it is analytically straightforward in the special case where the random matching is such that the types of any other two group members are statistically independent, given the member's own type. The morality profile is then a binomial distribution, and the utility function of a *Homo moralis* individual i is the expected value of i 's material payoff if, hypothetically, other members of the group would randomly and statistically independently switch to use i 's strategy with probability κ , which is then i 's degree of morality. At one end of the interval of such *Homo moralis*, $\kappa=0$, we find *Homo oeconomicus*; at the other end, $\kappa=1$, we find *Homo kantiansis*.

8. The general definition of *Homo moralis* is given in Alger and Weibull (2016).

Moreover, in large groups, the share of mutants in a mutant's group is, by the de Moivre–Laplace theorem, approximately normally distributed with mean value κ and variance $\kappa(1 - \kappa)/(n - 1)$. Hence, the share of other mutants is then almost deterministic and is equal to κ . A *Homo moralis* with degree of morality κ then acts (approximately) as if she hypothetically assumed that her behavior were to become, if not a “universal law,” then a “random law” applying to a randomly sampled share of size κ out of her group's other members.⁹

It is worth noting that the utility function of *Homo moralis* differs sharply from any utility function that only depends on the payoffs to all participants, such as altruism, inequity aversion, or a concern for social efficiency. We illustrate this by way of a simple example at the end of the third section.

Morality and ethics in connection with economics have been discussed at great length by many economists and philosophers, including Smith ([1759] 1976), Edgeworth (1881), Rawls (1971), Arrow (1973), Sen (1977), and Harsanyi (1980), to mention a few. But to the best of our knowledge, *Homo moralis* preferences have not been previously studied, or even known, with one exception. Bergstrom (1995) shows that evolutionary stability of strategies in interactions between siblings induces behavior that he calls “semi-Kantian,” which corresponds to $\kappa = 1/2$ in our equation (1).¹⁰

Kantian Morality and Economics

Economists' policy advice traditionally relies on models in which individuals have *Homo oeconomicus* preferences. What if economists' models instead were populated by the more general *Homo moralis*? In this chapter, we will merely scratch the surface by studying only a few examples.

9. This claim is not fully general and deserves further analysis, because even small perturbations of continuous (utility) functions may lead to “jumps” in behavior.

10. Bergstrom thus differs from us in studying stability of strategies rather than of utility functions. However, in Alger and Weibull (2013, corollary 5), we establish a link between these approaches by showing that *Homo moralis* equilibrium strategies are stable under strategy evolution. For a discussion of several ethical principles in relation to strategy evolution, see Bergstrom (2009).

Trust

There is variation across countries in the extent to which people are trusting, and trust is correlated with economic growth (Algan and Cahuc 2010).¹¹ In economics, the so-called *trust game* has been used extensively in controlled laboratory experiments as a way to measure trust and trustworthiness in different countries and cultures. This literature was pioneered by Berg, Dickhaut, and McCabe (1995) and has received a lot of attention among behavioral economists and experimentalists. The trust game is succinctly described by Cesarini et al. (2008, 3721):

Many mutually beneficial transactions involve an element of interpersonal trust and may fail to materialize in the absence of an expectation that trust will be reciprocated. The prevalence of trust in a society has therefore been assigned primacy in a number of domains, for instance empirical and theoretical studies of economic growth. In recent years, the trust game has emerged as a favorite instrument to elicit an individual's interpersonal trust and willingness to reciprocate trust. More generally, the game has been widely used to study cooperative behavior. In a trust game, an individual (the investor) decides how much money out of an initial endowment to send to another subject (the trustee). The sent amount is then multiplied by some factor, usually three, and the trustee decides how much of the money received to send back to the investor. The standard game-theoretic prediction for a single anonymous interaction between two purely self-interested individuals is for the investor to send nothing, rationally anticipating that the trustee will not reciprocate. Yet, experiments consistently show that cooperation flourishes in the trust game; the average investor sends a significant share of her endowment, and most trustees reciprocate.

What will *Homo moralis* do in such an interaction? Consider a situation in which two ex ante identical individuals are randomly paired. With equal chance, one of them is offered an endowment and an investment opportunity as described above. The other individual then has to act in the role of

11. A situation where trust is key is that of informal personal lending. In many developing countries, large fractions of the populations are still shut out from formal credit markets; see, for example, Kendall, Mylenko, and Ponce (2010). Then informal lending, in the form of not legally binding loans between individuals, can sometimes be enforced by the threat of future nonrenewal of lending (Ghosh and Ray 2016), social disapproval, or both. Evidence from laboratory experiments suggests that such informal lending may in fact even take place in one-shot interactions (Charness and Dufwenberg 2006). The trust game we analyze here can be interpreted as informal lending.

the trustee. A *strategy* for an individual in such a symmetric interaction then has two components. First, if given the endowment, what share $s \in [0, 1]$ of it is to be invested? Second, if not given the investment opportunity, what “payback rule” $p \in [0, 1]$ is to be used? Here, such a payback rule prescribes for any invested share $t \in [0, 1]$ chosen by the other party what share p of the gross return to pay back. Let $u(c)$ be an individual’s hedonic utility from own consumption c , and take this to represent the material payoff in our evolutionary framework. In the standard version of the trust game, the material payoff from using a strategy $x=(s, p)$ when the other individual uses strategy $y=(t, q)$ is then

$$\pi(x, y) = \frac{1}{2}u(1-s+3sq) + \frac{1}{2}u(3t-3tp). \quad (2)$$

In an interaction between two *Homo oeconomicus*, no party is trustworthy; they will choose $p=q=0$ for all $s, t > 0$. Thus, if each party knows the other’s type, no investment is made in equilibrium ($t=s=0$). The resulting expected material payoff to each party is $u(1)/2$, the probability of being given the initial endowment times the utility from keeping it. If instead both parties were *Homo kantiensis*, then they would each invest all the money if given the opportunity ($t=s=1$) and return half of the gross return (i.e., use payback rules p and q such that $p=q=0.5$). The resulting expected material payoff to each party is then $u(1.5)$, much higher than what *Homo oeconomicus* obtains.

Full morality is not necessary to induce full investment, however. For a pair of equally moral *Homo moralis*, full investment ($t=s=1$) obtains in equilibrium for any sufficiently high degree of morality, although as soon as morality is less than full ($\kappa < 1$), the trustee pays back less than half the gross returns from investment, in which case the trustee ends up being better off than the investor. As the degree of morality κ falls, the amount paid back decreases, and it eventually falls short of the amount originally invested, in which case the investor makes a material loss; nonetheless, morality makes the investor accept this loss and invest anyway, up to some point.¹² Indeed, for sufficiently low degrees of morality, the investor invests

12. To see this, note that the derivative of $U_\kappa(x, y)$ with respect to s , where $x=(s, p)$ and $y=(t, q)$, and evaluated when $t=s=1$, is positive even for $p < 1/3$ for $\kappa < 1$ large enough.

less than his full endowment, and eventually, when morality drops below a certain level, he invests nothing.

Public Goods

Many situations that are important for economic growth may be represented as situations in which people can make voluntary contributions to a public good, including the generation and dissemination of knowledge, and institution building. We examine the behavior of individuals in a community of n members, each of whom is in a position to make a voluntary contribution to a public good (the contribution may be monetary or in kind). A standard concern in economics is that free riding is enhanced as groups become larger, so our aim here is to analyze how group size affects the behavior of *Homo moralis*.

Suppose, then, that i obtains material payoff

$$\pi(x_i, y) = B\left(x_i + \sum_{j \neq i} y_j\right) - C(x_i) \quad (3)$$

if she makes the contribution x_i and the sum of the contributions from the other community members is Σy_j . Here B is a production function for the public good; and C a cost function for a contributing individual, representing forgone private consumption, income, or leisure. We take the marginal cost of making a contribution to be increasing, and the marginal benefit of the aggregate contribution to be decreasing.

Consider first the socially optimal individual contribution, x^* . With a conventional production function of the power form $B(X) = X^a$, where $0 < a < 1$, the necessary first-order condition for the sum of all members' material payoffs to be maximized,

$$nB'(nx^*) = C'(x^*), \quad (5)$$

implies that the socially optimal individual contribution x^* is increasing in n . By contrast, in a community of *Homo oeconomicus*, the first-order condition for the unique Nash equilibrium contribution, \hat{x}_0 , is

$$B'(n\hat{x}_0) = C'(\hat{x}_0), \quad (5)$$

which implies that in communities with more members, each individual contributes less. As a consequence, free riding—the tendency for people to under-provide public goods—is exacerbated when group size increases. The intuition is that if all contributions were to remain unchanged, then the marginal benefit from each contribution would fall. Thus, each individual will have a weaker incentive to contribute.

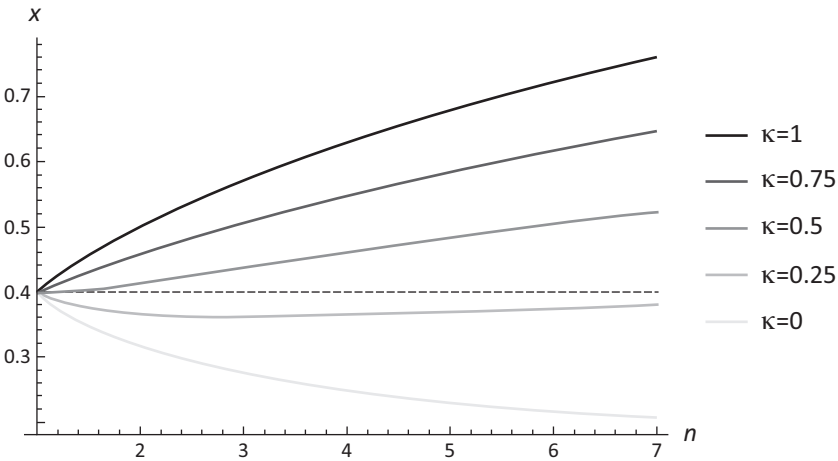


Figure 9.1

The unique Nash equilibrium contribution in the public-goods game for different degrees of morality

Suppose now instead that everyone in the community is a *Homo moralis* with the same degree of morality $\kappa \in [0, 1]$. Then their unique individual equilibrium contribution, \hat{x}_κ , can be shown to satisfy

$$[1 + (n - 1)\kappa] \cdot B'(n\hat{x}_\kappa) = C'(\hat{x}_\kappa), \tag{6}$$

For any positive degree of morality, group size has two counteracting effects on the individual contribution. The negative effect is, as before, due to the decreasing marginal productivity. The positive effect is that in larger groups, each individual’s contribution benefits a larger number of individuals. The “right thing to do,” as the group increases, is thus to increase one’s contribution. The positive effect may outweigh the negative.

To see this, consider again the conventional production function used above, and note that for purely Kantian individuals ($\kappa = 1$), the individual contribution always increases with n . For intermediate values of κ , the individual contribution decreases with n when small but increases with n when large. See figure 9.1, which shows the equilibrium contribution of *Homo moralis* with degree of morality κ as a function of community size n , with higher curves for higher degrees of morality (when $B(X) = \sqrt{X}$ and $C(x) = x^2$).

These predictions may potentially help explain observations made in laboratory experiments, in which group size sometimes has a positive effect and sometimes a negative effect on individual contributions (see Nosenzo, Quercia, and Sefton (2015) for a review).

Does the extent of free riding increase or decrease as group size increases? In the parametric specification used in figure 9.1, the individual contribution relative to the first-best contribution is

$$\frac{\hat{x}_\kappa}{x^*} = \left(\kappa + \frac{1-\kappa}{n} \right)^{2/3}, \quad (7)$$

a ratio that decreases as group size n increases (for any given degree of morality $\kappa < 1$).¹³ A smaller ratio indicates more free riding, so this equation shows that as morality κ increases, the effect of group size n on the extent of free riding declines.¹⁴ Moreover, the extent of free riding is bounded from below; as seen in (7), the ratio \hat{x}_κ/x^* exceeds $\kappa^{2/3}$ for all group sizes n . Hence, compared to the outcome under *Homo oeconomicus*, an important policy implication is that, when κ is positive, the contributions from *Homo moralis* decline less with group size and remain positive even in infinitely large groups.

Environmental Economics

According to World Bank president Jim Yong Kim, “If we don’t confront climate change, we won’t end poverty.”¹⁵ Some instruments have been proposed to help mitigate climate change, such as a carbon tax, regulation of production technologies, subsidies to public transportation, and support for R&D concerning environmentally friendly technologies for different forms of green energy. Determining the “right” carbon tax requires knowing how it will affect behavior and welfare. Here we briefly analyze the behavior of *Homo oeconomicus* and more generally, *Homo moralis*, in a standard model of consumption that has an external effect on the environment (Musgrave 1959, Arrow 1970). In this model, the group is taken to be so large that each individual’s impact on the group’s environment is negligible.

More specifically, there is a continuum of consumers, indexed $i \in I = [0, 1]$, and there are two consumption goods, goods 1 and 2, where good 1

13. Formally, $d(\hat{x}_\kappa/x^*)/dn < 0$ when $0 < \kappa \leq 1$.

14. Formally, $d^2(\hat{x}_\kappa/x^*)/(dn^2) > 0$.

15. See <http://www.worldbank.org/en/news/feature/2014/03/03/climate-change-affects-poorest-developing-countries>.

is environmentally neutral (that is, its consumption has no effect on the environment) and good 2 is environmentally harmful. Aggregate consumption of these goods are

$$X_1 = \int_I x_1(i) d\mu \quad \text{and} \quad X_2 = \int_I x_2(i) d\mu$$

where $x(i) = (x_1(i), x_2(i))$ is the consumption bundle of individual i , and μ is a density on I . Because all consumers are infinitesimally small, aggregate consumption is unaffected by any individual's personal consumption.

We take the material payoff to each individual i to be that individual's hedonic utility from own consumption, $x(i)$, and from the quality of the environment, which in turn depends on aggregate consumption, X_2 , of the environmentally harmful good. We write $u(x_1(i), x_2(i), X_2)$ for this hedonic utility and assume that it is increasing in consumption of each good and decreasing in aggregate consumption of the environmentally harmful good. Using good 1 as the numeraire, writing p for the price of good 2, and assuming that all individuals have the same income, a socially efficient consumption bundle, x^* , the same for all individuals i , satisfies

$$\frac{u_2(x_1^*, x_2^*, X_2^*)}{u_1(x_1^*, x_2^*, X_2^*)} = p - \frac{u_3(x_1^*, x_2^*, X_2^*)}{u_1(x_1^*, x_2^*, X_2^*)}, \quad (8)$$

where subscripts on the personal utility function denote partial derivatives. The marginal rate of substitution between the environmentally harmful and environmentally neutral goods should thus equal the relative price of the harmful good net of the marginal rate of substitution between the utility from the quality of the environment and the neutral good. In other words, social efficiency requires that, at given prices, consumers consume less of a good the more harmful it is to the environment.

By contrast, in a population consisting entirely of *Homo oeconomicus*, an (interior) equilibrium allocation in which everybody consumes the same bundle x^0 necessarily satisfies the first-order condition

$$\frac{u_2(x_1^0, x_2^0, X_2^0)}{u_1(x_1^0, x_2^0, X_2^0)} = p. \quad (9)$$

Under decreasing marginal utility of consumption, this means that *Homo oeconomicus*, not surprisingly, consumes more of the environmentally harmful good than required by social efficiency.

As observed above, for interactions in infinitely large groups, the utility function of an individual *Homo moralis* with degree of morality $\kappa \in [0,$

1] is the material payoff that would obtain if a share κ of the group would behave in the same way as the individual himself. In the present context, if an individual consumes the bundle $x=(x_1, x_2)$ and all the others consume some bundle $y=(y_1, y_2)$, then the utility to a *Homo moralis* with degree of morality κ would be

$$U_\kappa(x, y) = u(x_1, x_2, (1-\kappa)y_2 + \kappa x_2), \quad (10)$$

where, in this expression, we have normalized the total mass of individuals in the group (which could be a village, region, country, continent, or the whole world) to unity. In a group consisting entirely of *Homo moralis* with the same degree of morality κ , an (interior) equilibrium allocation, everybody consumes the same bundle x^κ , and this satisfies the first-order condition

$$\frac{u_2(x_1^\kappa, x_2^\kappa, x_2^\kappa)}{u_1(x_1^\kappa, x_2^\kappa, x_2^\kappa)} = p - k \cdot \frac{u_3(x_1^\kappa, x_2^\kappa, x_2^\kappa)}{u_1(x_1^\kappa, x_2^\kappa, x_2^\kappa)}. \quad (11)$$

Compared to *Homo oeconomicus*, for any positive degree of morality κ , each individual refrains somewhat from consuming the environmentally harmful good, although each individual—knowing that she is negligible—is fully aware that her own consumption has no effect on the overall quality of the environment! Hence, if people are in fact somewhat moral, then policy advice based on models inhabited by *Homo oeconomicus* may exaggerate the need for pecuniary incentives, such as carbon taxes. If people are more like *Homo moralis* with some positive degree of morality, then, in addition to some carbon taxes, it may be effective to provide individuals with information about how aggregate consumption (and production) creates carbon dioxide and what we know about how this affects the climate.¹⁶ By contrast, such information in this stylized example would have no effect at all on the behavior of *Homo oeconomicus*.¹⁷

16. Note that equations (8) and (9) are the special cases of (10) when $\kappa=0$ (*Homo oeconomicus*) and $\kappa=1$ (*Homo kantiansis*). Laffont (1975) considers these two extreme cases of self-interested individuals (our *Homo oeconomicus*) and “Kantian individuals” (our *Homo kantiansis*).

17. Note further that if good 2 does not cause any externality ($u_3=0$), then *Homo moralis* would behave precisely as the classical *Homo oeconomicus*; equation (10) would boil down to equation (9). For such goods, there is no “right thing to do,” and hence, morality has no bite.

Voting

Another class of situations in which *Homo moralis* may make a difference is collective decision-making by voting. By and large, countries with more developed economies tend to have more democratic political systems (see, e.g., Persson and Tabellini (2006) and Acemoglu et al. (2014)). For democracy to work, it is important that citizens participate in elections, committee work, and related activities; it is still much debated in economics and political science why and how people vote. As has been pointed out by economists, high participation rates in large elections appear incompatible with rational *Homo oeconomicus* behavior. The reason being that the act of voting usually has some personal cost, say, lost income or leisure, and this cost easily outweighs the expected benefit to the individual of participating in the election, because the probability of being pivotal is virtually nil. This is the well-known *voters' paradox*. Despite this, the turnout in general and local elections in many countries is often impressive. So what then motivates people to participate in elections? Can *Homo moralis* provide an explanation?

A closely related and arguably equally important issue is participation and voting in committees, such as parliamentary bodies, company boards, court juries, and central bank boards. As shown by Austen-Smith and Banks (1996), when committee members have private information and are *Homo oeconomicus*, then voting may fail to aggregate information efficiently, even when the members have the same preferences. This observation challenges the so-called Condorcet jury theorem (Condorcet 1785), which states that democracy in the form of majority rule in such situations is a great institution, because it implies that the right decision is almost always taken if the electorate is large enough. How would *Homo moralis* vote in such committees?

Other Social Preferences

Theoretical work on the evolutionary foundations of human motivation provides insights about potential *ultimate* causes of human behavior—the forces in the environment that have shaped our preferences, not only for the foods that contain the nutrition that we need to survive but also for behaviors in social interactions. This line of research is complementary to behavioral economics, the branch of economics that investigates the explanatory power of richer motivations than mere self-interest. In the language of evolutionary biology, the focus in behavioral economics is

on the *proximate* causes of observed human behaviors—the neurological, hormonal, and psychological mechanisms and triggers that induce us to behave in certain ways. Here we briefly discuss how *Homo moralis* preferences compare with those considered in this literature, which is inspired by research in psychology and sociology.

In the 1970s and 1980s, altruistic preferences were proposed to explain intra-family transfers, transfers to the poor, and contributions to public goods (Becker 1974, 1976; Andreoni 1988; Lindbeck and Weibull 1988). However, altruism turned out to be insufficient to explain the data, and “warm glow” was then proposed to enhance the understanding of voluntary contributions to public goods (Andreoni 1990). In the 1990s, inequity aversion, or a preference for fairness, was introduced by Fehr and Schmidt (1999) as an explanation for why people have a tendency to turn down low offers in the ultimatum bargaining game (Güth, Schmittberger, and Schwarze 1982). Still other forms of human motivation that have been proposed, and sometimes tested, include conformity (Bernheim 1994), conditional altruism (Levine 1998), identity (Akerlof and Kranton 2000), and honesty and truth telling (Alger and Ma 2003; Alger and Renault 2006; Demichelis and Weibull 2008).

Although conceptually very different from *Homo moralis*, these preferences would be compatible with evolutionary stability if they gave rise to the same equilibrium behaviors as those of *Homo moralis*.¹⁸ For what class of material payoff functions such behavioral equivalence obtains remains to be analyzed. Here we limit ourselves to pointing out that *Homo moralis* preferences sometimes give rise to radically different behaviors compared to preferences that may appear to be similar. For example, consider altruistic preferences. An altruistic individual's preferences are usually represented as a utility function that attaches unit weight to the individual's own material payoff and a positive weight, less than 1, to other individuals' material payoffs. An altruist hence internalizes some of the external effects of her behavior on others. Let the latter weight be denoted α , the individual's *degree of altruism* toward the other party.¹⁹ For some material payoff func-

18. However, the preferences of *Homo moralis* are the only ones that are evolutionarily stable in the whole class of interactions analyzed in Alger and Weibull (2013, 2016).

19. For $n=2$, an altruist's utility is $u_\alpha(x, y) = \pi(x, y) + \alpha\pi(y, x)$. Note that this function can also be interpreted as the individual having a concern for efficiency, because it is a monotone transformation of $v_\alpha(x, y) = \pi(x, y) + \frac{\alpha}{1-\alpha} [\pi(x, y) + \pi(y, x)]$.

tions, an altruist with degree of altruism α behaves exactly like *Homo moralis* with a degree of morality $\kappa = \alpha$ (see Alger and Weibull 2013). Hence, in some interactions, one cannot discriminate between moralism and altruism as explanations for observed behavior. However, the two classes of preferences are conceptually quite distinct and induce radically different behaviors in some interactions. This difference is particularly striking in interactions with many participants and in coordination problems among few or many participants.

To illustrate the first case, consider again the environmental economics and the public goods examples. In the environmental example, morality induced consumers to reduce their consumption of the harmful good, even though the effect of each individual's consumption was negligible. In the public goods example, as the number of participants tends to infinity, the individual contribution to the public good tends to a positive amount for any positive degree of morality. By contrast, Andreoni (1988) has shown that in a population of altruists, the proportion of individuals who make positive donations shrinks to zero as the number of individuals grows infinitely large, because each individual donation then has a negligible effect on the total value of the public good. There is thus a sharp distinction between morality and altruism when groups are large. Even if an individual is highly altruistic and cares about the consequences of her behavior for others, she will behave very much like *Homo oeconomicus* if her impact is marginal. By contrast, *Homo moralis* cares directly about his own behavior, beyond the effects that this behavior has on his own material payoff, and this consideration for "the right thing to do" makes him behave differently from both selfish and altruistic individuals in these situations.

This observation may have important implications for other policy issues as well, such as tax compliance. It has been noted by some economists (see Sandmo 2005), that less tax evasion appears to occur in certain countries than would be compatible with *Homo oeconomicus*'s behavior. The risk of being caught is often small and the penalties mild, so maximization of expected personal utility would suggest rampant tax evasion. So why do people in those countries, and perhaps many in other countries, not evade taxes more? Because the marginal effect of any change in an individual's tax payment is, with few exceptions, negligible, prosocial preferences such as altruism or inequity aversion may fail to explain why individuals evade taxes. However, as suggested by the analysis above,

Homo moralis may supply an explanation, because a *Homo moralis* may, to a certain extent, prefer to pay taxes, since she cares about the moral quality of her actions.

Let us now turn to the second situation in which *Homo moralis* preferences give rise to radically different behaviors compared to altruism, namely, coordination problems. Consider an example from Alger and Weibull (2013), a simple 2×2 coordination game in terms of material payoffs:

| | | | |
|---|-----|-----|------|
| | A | B | |
| A | 2,2 | 0,0 | (12) |
| B | 0,0 | 1,1 | |

When individuals pair up to play this game, two alternative potential societal “conventions” are available: Either both parties take action A, or both parties take action B. Clearly, the first convention is Pareto superior to the second. However, under each convention, *Homo oeconomicus* has no incentive to unilaterally deviate. Granted that a sufficiently large population share act according to the going convention, an individual deviator would lose material payoff and, in addition, inflict a payoff loss on the unfortunate opponent.²⁰ Therefore, an altruist would also stick to the going convention, even if this happened to be the socially inferior convention to always take action B. But not so a *Homo moralis* of high enough degree of morality. For suppose a *Homo kantiensis* were to visit a country where (by and large) every citizen takes action B in every encounter, and suppose that the visitor is indistinguishable from a citizen. Then *Homo kantiensis* would take action A in each encounter, because this would be “the right thing to do” if upheld as a universal law of conduct.²¹ This moralistic visitor will earn material payoff zero in each encounter, and so will the unfortunate citizens who meet him. The citizens would very much wish that the visitor instead had been a *Homo oeconomicus* or an altruist.

20. These are strict Nash equilibria in terms of material payoffs. The game also has a mixed equilibrium, in which each individual plays A with probability $1/3$. However, this equilibrium is unstable in all plausible population dynamics. See Young (1993) and Myerson and Weibull (2015) for formal models of stable conventions in large populations.

21. Indeed, to take action A is optimal for all *Homo moralis* individuals with degree of morality $\kappa \geq 1/3$.

A final point before concluding. Some researchers have developed models in which individuals care about norms, have a concern for their image (in the eyes of others and perhaps also in their own eyes), or a desire to avoid social stigma (Lindbeck, Nyberg, and Weibull 1999; Brekke, Kverndokk, and Nyborg 2003; Bénabou and Tirole 2006; Ellingsen and Johannesson 2008; Huck, Kübler, and Weibull 2012). In these models, individuals are assumed to have a baseline intrinsic wish to “behave well” and in addition a wish to be viewed favorably by others, image concerns that may strengthen the wish to behave well (Bénabou, Falk, and Tirole 2018). Evidently, we humans are very complex creatures, and our behavior is most likely driven by many motives (what biologists would call “proximate causes” for our actions). Biologists distinguish such proximate causes from ultimate causes, by which is meant the reasons we exist in the evolutionary race. Our derivation of *Homo moralis* was based entirely on such ultimate causes. A closer examination of relations between proximate and ultimate causes in human motivation is an avenue for future research. Eventually, evolutionary theory may help close the open-endedness of behavioral economics by providing testable predictions regarding which preferences are more likely to be sustained.

Conclusion

In this chapter, we have discussed the evolutionary foundations for human motivation, how evolution favors the class of *Homo moralis* preferences, and the implications for economics and policy of such preferences compared to other preferences. We have presented the following main points:

1. Economics possesses powerful analytical tools that enable positive and normative analyses of a wide range of social and economic phenomena. These tools should not be abandoned but instead brought to more general use.
2. The conventional assumption among economists, since the days of Adam Smith's ([1776] 1976) *Wealth of Nations*, is that economic agents are purely self-interested and focused on their own consumption. Yet behavioral and experimental economics, insights from the other social and behavioral sciences, everyday observation, and introspection suggest that human motivation is much more complex, sometimes systematically deviating from narrow self-interest.

3. First principles in evolutionary biology, formalized in terms of evolutionary stability along the lines of Maynard Smith and Price (1973) suggest that, in our simple model framework, evolution favors human motivation in the form of *Homo moralis*, a generalization of *Homo oeconomicus* that allows for varying degrees of morality along with self-interest.
4. Applying the powerful analytical tools of economics to *Homo moralis* results in new predictions and policy recommendations. In particular, because *Homo moralis* is not only motivated by her material gains and losses, policy based on *Homo oeconomicus* may lead to exaggerated use of pecuniary incentives, such as distortionary taxes. If people do have a natural inclination for moral concerns, it may be more effective to provide the public with information about the consequences of our actions, for ourselves and others.

Our results being purely theoretical, empirical and experimental work will be necessary to determine the empirical validity of *Homo moralis*. To this end, further theoretical analysis is also needed: Even though we have here examined the behavior of *Homo moralis* in some common situations, we have only scratched the surface. Moreover, many fundamental questions have not been addressed at all. In particular, one fundamental issue that we have not (yet) addressed is welfare. For economic and social policy, this is a most important and philosophically nontrivial issue, especially when individuals have social preferences. If individuals have *Homo moralis* preferences, perhaps idiosyncratic degrees of morality, should welfare then be defined in terms of the material payoffs or in terms of individuals' utility functions?

This philosophically and methodologically difficult issue may be related to that addressed by John Harsanyi in two wonderful essays that deal with game theory, utilitarianism and ethics (Harsanyi (1980, 1992)). In these essays, he advocates what he calls "rule utilitarianism," an approach we find also appealing for *Homo moralis*. Harsanyi distinguishes between an individual's "personal preferences" and his or her "moral preferences." He advocates that, when defining welfare in a society, one should only consider personal preferences. When individuals' preferences can be represented by an additive utility function, where one term can be taken to represent "personal utility," Harsanyi argues that welfare should be defined as the sum of all individuals' expected personal utilities, behind the veil of ignorance as to

what societal position each individual will end up in. This appears to be in line with *Homo moralis*. If we take the material payoff function to represent personal utility, then welfare in a society consisting of *Homo moralis* individuals (each with his or her degree of morality) should be defined simply as the sum of their expected material payoffs, just as in ordinary utilitarian welfare theory.

To wit, suppose a parent has one selfish and one altruistic child, and has a cake to divide between them. Suppose also that both children have the same hedonic utility from consumption, and that this is increasing in the amount consumed, with decreasing marginal utility.²² Should the parent give a bigger slice to the selfish child, thus maximizing the sum of their altruistic and selfish utilities, or should the parent give them equally large slices, thus maximizing the sum of only their hedonic utilities? The second alternative undoubtedly seems more appealing. The same could be said with one selfish and one spiteful child; taking into account both children's total utility, a bigger slice should be given to the spiteful child, but equal division is, arguably, more reasonable. By contrast, if one child is selfish, and the other instead is inequity averse or a *Homo moralis* (with any degree of morality), it makes no difference if the parent considers the children's total or hedonic (personal) utilities; in every case, their joint welfare is maximized by equal division. Further study of the welfare economics of *Homo moralis* and other social preferences is a topic for future research.

A final point we make concerns the status of economics as a discipline, in the eyes of the general public and among the other behavioral and social sciences. Conventional economics textbooks may give the false impression that selfishness is part of economic rationality (see the discussion in Rubinstein (2006) and the references therein). This misreading of conventional economics probably hurts the reputation of economists. If economists would instead use partly morally motivated agents, such as *Homo moralis*, then such misunderstandings could be avoided, and the critique would fall flat to the ground. The economist's analysis would then not be prejudiced in favor of either selfishness or morality. Instead it would allow for the whole spectrum of intermediate degrees of morality, spanning from pure self-interest to pure Kantian morality.

22. This example is due to Peter Diamond, discussed in a conversation many years ago with one of the authors.

References

- Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A. Robinson. 2014. "Democracy Does Cause Growth." NBER Working Paper 20004, National Bureau of Economic Research, Cambridge, MA.
- Akerlof, George, and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115 (3): 715–753.
- Alexander, Richard D. 1987. *The Biology of Moral Systems*. New York: Aldine De Gruyter.
- Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review* 100 (5): 2060–2092.
- Alger, Ingela, and Albert Ma. 2003. "Moral Hazard, Insurance, and Some Collusion." *Journal of Economic Behavior and Organization* 50 (2): 225–247.
- Alger, Ingela, and Régis Renault. 2006. "Screening Ethics When Honest Agents Care about Fairness." *International Economic Review* 47 (1): 59–85.
- Alger, Ingela, and Jörgen W. Weibull. 2010. "Kinship, Incentives, and Evolution." *American Economic Review* 100 (4): 1725–1758.
- Alger, Ingela, and Jörgen W. Weibull. 2013. "Homo Moralistic—Preference Evolution under Incomplete Information and Assortativity." *Econometrica* 81 (6): 2269–2302.
- Alger, Ingela, and Jörgen W. Weibull. 2016. "Evolution and Kantian Morality." *Games and Economic Behavior* 98: 55–67.
- Andreoni, James. 1988. "Privately Provided Public Goods in a Large Economy: The Limits of Altruism." *Journal of Public Economics* 35 (1): 57–73.
- Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100 (401): 464–477.
- Arrow, Kenneth. 1970. "Political and Economic Evaluation of Social Effects and Externalities." In *The Analysis of Public Output*, edited by Julius Margolis, 1–23. New York: Columbia University Press.
- Arrow, Kenneth. 1973. "Social Responsibility and Economic Efficiency." *Public Policy* 21: 303–317.
- Austen-Smith, David, and Jeffrey S. Banks. 1996. "Information Aggregation, Rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90 (1): 34–45.
- Banerjee, Abhijit, and Jörgen W. Weibull. 1995. "Evolutionary Selection and Rational Behavior." In *Learning and Rationality in Economics*, edited by Alan Kirman and Mark Salmon, 343–363. Oxford: Basil Blackwell.
- Becker, Gary S. 1974. "A Theory of Social Interaction." *Journal of Political Economy* 82 (6): 1063–1093.

- Becker, Gary S. 1976. "Altruism, Egoism, and Genetic Fitness: Economics and Sociobiology." *Journal of Economic Literature* 14 (3): 817–826.
- Bénabou, Ronald, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–1678.
- Bénabou, Ronald, Armin Falk, and Jean Tirole. 2018. "Narratives, Imperatives and Moral Reasoning." Mimeo, Toulouse School of Economics, France.
- Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10 (1): 122–142.
- Bergstrom, Theodore C. 1995. "On the Evolution of Altruistic Ethical Rules for Siblings." *American Economic Review* 85 (1): 58–81.
- Bergstrom, Theodore C. 2003. "The Algebra of Assortative Encounters and the Evolution of Cooperation." *International Game Theory Review* 5 (3): 211–228.
- Bergstrom, Theodore C. 2009. "Ethics, Evolution, and Games among Neighbors." Working Paper, University of California, Santa Barbara.
- Bergstrom, Theodore C. 2012. "Models of Assortative Matching." Working Paper, University of California, Santa Barbara.
- Bernheim, B. Douglas. 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–877.
- Brekke, Kjell A., Snorre Kverndokk, and Karine Nyborg. 2003. "An Economic Model of Moral Motivation." *Journal of Public Economics* 87 (9): 1967–1983.
- Cesarini, David, Christopher T. Dawes, James H. Fowler, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2008. "Heritability of Cooperative Behavior in the Trust Game." *Proceedings of the National Academy of Sciences* 105 (10): 3721–3726.
- Charness, Gary, and Martin Dufwenberg. 2006. "Promises and Partnership." *Econometrica* 74 (6): 1579–1601.
- Condorcet, Nicolas de. 1785. *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. Paris: L'Imprimerie Royale.
- Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities and Segregation." *Econometrica* 77 (4): 1003–1045.
- Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2010. "Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation." *Proceedings of the National Academy of Sciences* 107 (11): 4857–4861.
- Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection*. London: John Murray.

- Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford University Press.
- Dekel, Eddie, Jeffrey C. Ely, and Okan Yilankaya. 2007. "Evolution of Preferences." *Review of Economic Studies* 74 (3): 685–704.
- Demichelis, Stefano, and Jörgen W. Weibull. 2008. "Language, Meaning, and Games: A Model of Communication, Coordination, and Equilibrium." *American Economic Review* 98 (4): 1292–1311.
- Edgeworth, Francis Y. 1881. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London: Kegan Paul.
- Ellingsen, Tore, and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (3): 990–1008.
- Fehr, Ernst, and Klaus Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (3): 817–868.
- 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.
- Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.
- Ghosh, Parikshit, and Debraj Ray. 2016. "Information and Enforcement in Informal Credit Markets." *Economica* 83 (329): 59–90.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization* 3 (4): 367–388.
- Güth, Werner, and Menahem E. Yaari. 1992. "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach." In *Explaining Process and Change: Approaches to Evolutionary Economics*, edited by Ulrich Witt, 23–24. Ann Arbor: University of Michigan Press.
- Hamilton, William D. 1964. "The Genetical Evolution of Social Behaviour." *Journal of Theoretical Biology* 7 (1): 1–16.
- Harsanyi, John C. 1980. "Rule Utilitarianism, Rights, Obligations and the Theory of Rational Behavior." *Theory and Decision* 12 (2): 115–133.
- Harsanyi, John C. 1992. "Game and Decision Theoretic Models in Economics." In *Handbook of Game Theory with Economic Application*, volume 1, edited by Robert J. Aumann and Sergiu Hart, 669–707. Amsterdam: North-Holland.
- Heifetz, Aviad, Chris Shannon, and Yossi Spiegel. 2007. "What to Maximize If You Must." *Journal of Economic Theory* 133 (1): 31–57.

Huck, Steffen, Dorothea Kübler, and Jörgen W. Weibull. 2012. "Social Norms and Economic Incentives in Firms." *Journal of Economic Behavior and Organization* 83 (2): 173–185.

Kant, Immanuel. [1785] 2002. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven, CT: Yale University Press.

Kendall, Jake, Nataliya Mylenko, and Alejandro Ponce. 2010. "Measuring Financial Access around the World." World Bank Policy Research Working Paper 5253, World Bank, Washington, DC.

Laffont, Jean-Jacques. 1975. "Macroeconomic Constraints, Economic Efficiency and Ethics: An Introduction to Kantian Economics." *Economica* 42 (168): 430–437.

Levine, David K. 1998. "Modelling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics* 1 (3): 593–622.

Lindbeck, Assar, and Jörgen W. Weibull. 1988. "Altruism and Time Consistency: The Economics of *Fait Accompli*." *Journal of Political Economy* 96 (6): 1165–1182.

Lindbeck, Assar, Sten Nyberg, and Jörgen W. Weibull. 1999. "Social Norms and Economic Incentives in the Welfare State." *Quarterly Journal of Economics* 114 (1): 1–35.

Maynard Smith, John, and George R. Price. 1973. "The Logic of Animal Conflict." *Nature* 246 (November): 15–18.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415–444.

Musgrave, Richard A. 1959. *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill.

Myerson, Roger, and Jörgen W. Weibull. 2015. "Tenable Strategy Blocks and Settled Equilibria." *Econometrica* 83 (3): 943–976.

Nosenzo, Daniele, Simone Quercia, and Martin Sefton. 2015. "Cooperation in Small Groups: The Effect of Group Size." *Experimental Economics* 18 (1): 4–14.

Ok, Efe A., and Fernando Vega-Redondo. 2001. "On the Evolution of Individualistic Preferences: An Incomplete Information Scenario." *Journal of Economic Theory* 97 (2): 231–254.

Persson, Torsten, and Guido Tabellini. 2006. "Democracy and Development: The Devil in the Details." *American Economic Review* 96 (2): 319–324.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rousset, François. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton, NJ: Princeton University Press.

Rubinstein, Ariel. 2006. "A Sceptic's Comment on the Study of Economics." *Economic Journal* 116: C1–C9.

Ruef, Martin, Howard E. Aldrich, and Nancy M. Carter. 2003. "The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs." *American Sociological Review* 68 (2): 195–222.

Sandmo, Agnar. 2005. "The Theory of Tax Evasion: A Retrospective View." *National Tax Journal* 58 (4): 643–663.

Sen, Amartya K. 1977. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy & Public Affairs* 6 (4): 317–344.

Smith, Adam. [1759] 1976. *The Theory of Moral Sentiments*. Edited by David D. Raphael and Alec L. Macfie. Oxford: Oxford University Press.

Smith, Adam. [1776] 1976. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edited by William B. Todd. Oxford: Oxford University Press.

Wright, Sewall. 1931. "Evolution in Mendelian Populations." *Genetics* 16 (2): 97–159.

Young, H. Peyton. 1993. "The Evolution of Conventions." *Econometrica* 61 (1): 57–84.

This is a section of [doi:10.7551/mitpress/11130.001.0001](https://doi.org/10.7551/mitpress/11130.001.0001)

The State of Economics, the State of the World

Edited by: Kaushik Basu, David Rosenblatt,
Claudia Sepúlveda

Citation:

The State of Economics, the State of the World

Edited by: Kaushik Basu, David Rosenblatt, Claudia Sepúlveda

DOI: 10.7551/mitpress/11130.001.0001

ISBN (electronic): 9780262353472

Publisher: The MIT Press

Published: 2020



The MIT Press



This work is available under the Creative Commons Attribution—NonCommercial—NoDerivatives 3.0 IGO license (CC BY-NC-ND 3.0 IGO) <http://creativecommons.org/licenses/by-nc-nd/3.0/igo>.

Some rights reserved

The findings, interpretations, and conclusions expressed in this work are those of the authors and do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy, completeness, or currency of the data included in this work and does not assume responsibility for any errors, omissions, or discrepancies in the information, or liability with respect to the use of or failure to use the information, methods, processes, or conclusions set forth. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Nothing herein shall constitute or be construed or considered to be a limitation upon or waiver of the privileges and immunities of The World Bank, all of which are specifically reserved.

Attribution—Please cite the work as follows: The World Bank. 2019. *The state of economics, the state of the world* / edited by Kaushik Basu, Claudia Sepulveda, and David Rosenblatt. Published by MIT Press. © World Bank. License: Creative Commons Attribution—NonCommercial—NoDerivatives 3.0 IGO (CC BY-NC-ND 3.0 IGO).

Third-party content—The World Bank does not necessarily own each component of the content contained within the work. The World Bank therefore does not warrant that the use of any third-party-owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to re-use a component of the work, it is your responsibility to determine whether permission is needed for that re-use and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures, or images.

All queries on rights and licenses should be addressed to the Publishing and Knowledge Division, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: pubrights@worldbank.org.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Basu, Kaushik, editor. | Sepúlveda, Claudia Paz, 1969– editor. | Rosenblatt, David, editor.

Title: *The state of economics, the state of the world* / edited by Kaushik Basu, Claudia Sepulveda, and David Rosenblatt.

Description: Cambridge, MA : MIT Press, [2019] | Includes bibliographical references and index.

Identifiers: LCCN 2018046336 | ISBN 9780262039994 (hardcover : alk. paper)

Subjects: LCSH: Economic development. | Information technology—Economic aspects. | Monetary policy. | Social change.

Classification: LCC HD82 .S8223 2019 | DDC 330.1—dc23

LC record available at <https://lcn.loc.gov/2018046336>

10 9 8 7 6 5 4 3 2 1