THE
CRITERION
JOURNAL ON INNOVATION

VOL. 4 2019

Fairness

Ken Binmore[*]

I am finishing *Crooked Thinking or Straight Talk? Epicurus Shows the Way*, a book written for oddballs like myself, who like to do their own thinking for themselves without following the herd wherever it may go. Its theme is that modern philosophical ideas with a science base can help us structure the way we think about ourselves and our society in a practical way that is a million miles from the pretentious drivel offered by academic philosophers of the old school. This essay previews what my forthcoming book will say about fairness.

I. Evolutionary Ethics

Traditional philosophers don't think evolution is relevant to morality. Sometimes they even warn against the heresy represented by evolutionary ethics. So where do they think moral principles come from?

Theologians say they come from God. Various metaphysical substitutes for a God, such as Practical Reason, Moral Intuition, General Will, or Natural Law are popular with more secular philosophers. Immanuel Kant thought that he could deduce his categorical imperative from Rationality alone. According to him, all Rational Beings will therefore honor the same moral code.

Followers of Epicurus have little patience with the kind of empty playing with words in such fairy stories. We don't believe that morality somehow exists in some Platonic limbo independently of the physical world. We think that the moral norms found by anthropologists in all societies everywhere are products of the evolutionary history—both biological and social—of our species. They aren't even unique to Homo sapiens. The sharing of blood among vampire bats is the most exotic example, but chimps even follow

[*] Professor Emeritus, University College London. Email: k.binmore@ucl.ac.uk.

humans in operating norms that sometimes differ between different societies as a consequence of their differing social histories.

Anthropology is helpful in sorting out what is universal in the human species from what differs from one society to another. It is surely no accident that Kalahari bushmen, African pygmies, Andaman Islanders, Greenland Eskimos, Australian aborigines, Paraguayan Indians, and Siberian nomads and other pure hunter-gatherer societies that survived into the twentieth century all operated social contracts in which food, especially meat, was shared on a markedly egalitarian basis. But such small societies lived in a wide variety of diverse environments. What their social contracts have in common is therefore presumably biologically determined, leaving their many differences to be explained by social evolution. But even when such universal traits exist, they aren't absolute—they might have been different if evolution had gifted us with a different set of genes.

Sore thumbs. Traditionalists respond that evolution can't be responsible for morality, because Nature is red in tooth and claw. How can the survival of the fittest be compatible with loving your neighbor?

But only in the golden age of the poets can people be relied upon to emulate the Good Samaritan in loving strangers as they love themselves. Of course evolution won't generate the kind of saintly behavior that traditionalists identify with morality.

But we don't have to accept their criteria for what counts as moral. Evolution didn't shape our minds to operate the utopian fantasies invented by metaphysicians; it created the moral rules that we actually use in real life. The moral rules with which this essay is concerned evolved to solve everyday coordination problems.

The sort of coordination problems I have in mind are those that we commonly solve without thought or discussion, usually so smoothly and effortlessly that we don't even notice that there is a coordination problem to be solved. Who goes through that door first? How long does Alice get to speak before it is Bob's turn?

Who moves how much in a narrow corridor when a fat lady burdened with shopping passes a teenage boy with a ring through his nose? Who should take how much of a popular dish of which there isn't enough to go around? Who gives way to whom when cars are maneuvering in heavy traffic? Who gets that parking space? Whose turn is it to wash the dishes tonight? These are picayune problems, but if conflict arose every time they needed to be solved, our societies would fall apart.

Most people are surprised at the suggestion that there might be something problematic about how two people pass each other in the corridor. When interacting with people from our own culture, we commonly solve such coordination problems so effortlessly that we don't even think of them as problems. Our moral programming then runs well below the level of consciousness, like our internal routines for driving cars or tying shoelaces. As with Molière's Monsieur Jourdain, who was delighted to discover that he had been speaking prose all his life, we are moral in small-scale situations without knowing that we are moral.

Just as we only take note of a thumb when it is sore, we tend to notice moral rules only when attempts are made to apply them in situations for which they are ill-adapted. We are then in the same position as Konrad Lorenz when he observed a totally inexperienced baby jackdaw go through all the motions of taking a bath when placed on a marble-topped table. By triggering such instinctive behavior under pathological circumstances, Lorenz learned a great deal about what is instinctive and what is not when a bird takes a bath. But this vital information is gained only by avoiding the mistake of supposing that bath-taking behavior confers some evolutionary advantage on birds placed on marble-topped tables.

Similarly, we can learn a lot about the mechanics of moral norms by triggering them under pathological circumstances—but only if we don't make the mistake of supposing that the moral rules are adapted to the coordination problems they fail to solve. However, it is precisely from such sore-thumb situations that I think traditional moralists unconsciously distil their ethical principles. They discuss these situations endlessly, because our failure to coordinate successfully brings them forcefully to our attention.

Why did fairness norms evolve? The Epicurean answer is that they provide a convention that allowed us to coordinate on a compromise equilibrium in the Sharing Game played by our prehuman ancestors. Why did evolution care about sharing? Because sharing is a means of insuring against hunger. Why didn't the big guys grab the lot? Because of reciprocal altruism in the repeated Sharing Game.

We are by no means the only species that shares food, but each species has its own way of sharing. Vampire bats have their way of sharing and we have ours. We call our method of sharing fairness, and use it these days for a lot more than simply sharing food.

How do fairness norms work? Writers of footnotes to Plato think that science is helpless in the face of such questions, but this essay offers a possible answer. It is doubtless inadequate or plain wrong in places, but its point is to make it clear that we needn't heed the traditionalists when they tell us that evolutionary ethics is an absurd impossibility. The answer offered is very unlikely to be the final word on the subject, but it shows that evolutionary ethics is neither absurd nor impossible—merely unspectacular.

As Epicurus was perhaps first to appreciate, justice isn't some grand notion like the laws of physics, it is merely a device that evolved to solve equilibrium selection problems in our game of life. It is basically for this reason that it is unacceptable to traditional moral philosophers. They want justice to be a substitute for power rather than a means of balancing power. They are determined not to understand that it is the reciprocity principle that keeps our urge to dominate others in check, and not our respect for their invented moral principles.

Fair social contracts? Fairness is a social device washed up on our evolutionary beach for use in solving small-scale coordination problems. Can we adapt it for use in constitutional design to create a just society? I think it worth a try—but we won't get anywhere in seeking to apply our current fairness norms to large-scale problems if we don't understand how they currently work in solving small-scale problems. So this is our first objective.

II. The Golden Rule

The Golden Rule—do as you would be done by—seems to be a universal moral principle. A list of sages who endorse the Golden Rule would be endless. But here is what the superstars have to say:

> *Zoroaster*: That nature alone is good that refrains from doing to another what is not good for itself.
> *Buddha*: Hurt not others in ways that you yourself would find hurtful.
> *Confucius*: Do not unto others what you would not have them do unto you.
> *Hillel*: Do not to others that which you would not have others do to you.
> *Jesus*: Do to others whatever you would have them do to you.
> *Mohammed*: As you would have people do to you, do to them; and what you dislike to be done to you, don't do to them.

The Golden Rule captures something of the reciprocity principle (which I discuss in a previous chapter of my forthcoming book), but it is also a fairness criterion. As such, it has been much criticized because it fails to recognize that different people like different things. For example, it may be that Alice likes being shaken awake before dawn for a cold shower and a ten-mile run. Bob prefers a gentle awakening at a late hour with a cup of coffee and a newspaper. So Alice doesn't want Bob to do to her what he would like her to do for him. Bob would like it even less if Alice were to do to him what she would like him to do for her.

Empathetic preferences. The sages who argued for the Golden Rule would probably have regarded such objections as nitpicking. They would perhaps have said that they were taking for granted that the Golden Rule needs to be modified when people have different tastes. The qualified version is simply: Do as you would be done by—if you were the person to whom something is to be done.

If Alice wants to implement this qualified Golden Rule, she needs to be equipped with the empathetic preferences we meet when discussing utilitarianism. She not only needs to be able to put herself in Bob's position to see things from his point of view—she needs to be able to compare how she feels herself with how she would feel if she were Bob.

This seems so obvious in retrospect that it is hard to understand how moral philosophers could have overlooked the relevance of empathetic preferences for so long. Apart from anything else, speaking explicitly of empathetic preferences forces us to ask: How come people from the same culture are broadly in agreement on how to compare each other's welfare, so that Alice and Bob can be modeled as having the *same* empathetic preferences? We can dodge this question temporarily by allowing a metaphysical ideal observer called Carol to make fairness decisions on behalf of Alice and Bob. But we can't allow fanciful notions from the golden age of the poets in an evolutionary approach. We have to dispense with Carol and try to understand how the social consensus that she supposedly represents might evolve.

This creates some difficulties in language, because traditional discussions always correspond to adopting Carol's viewpoint. So weights in a traditional utilitarian sum are all taken to be equal because the utils to be summed can be identified with Carol's empathetic utils. Similarly, in traditional egalitarian discussions, Alice and Bob gain equally because the utils to be equalized are Carol's empathetic utils. In her absence, we only have Alice and Bob's personal utility scales to work with. It is then necessary to introduce weights into a utilitarian sum. In the egalitarian case, Alice and Bob's utility gains are only proportional instead of equal.

To tackle the social consensus problem and other questions left hanging in the air by the traditional approach to the Golden Rule, we need to think harder about how the Golden Rule works in practice.

A. The Original Position

The original position was invented in the 1950s by John Rawls—who is justly regarded as the leading moral philosopher of the last century.[1] He describes it as an operationalization of the categorical imperative. John Harsanyi proposed the same idea independently at about the same time. Both were Kantians interested in using the idea to characterize an ideally fair society.

[1] John Rawls, A Theory of Justice (Harvard Univ. Press rev. ed. 1999) (1971). I owe a debt of gratitude to Rawls for encouraging me to work on the ideas outlined in this essay. If everybody were like John Rawls, we wouldn't need to worry so much about how morality works.

It is ironic that Harsanyi was thereby led to argue for utilitarianism, and Rawls—who didn't approve of utilitarianism at all—to argue for a kind of egalitarianism.

I see the original position as a device we use in ordinary life to implement the Golden Rule in the qualified form that takes account of the fact that Alice and Bob may have different tastes. So we need to put aside for the moment its possible role in designing a fair constitution. We shall just talk about Alice and Bob solving some kind of sharing problem.

We also need to abandon the Kantian perspective of both Harsanyi and Rawls in favor of an evolutionary approach. It is true that most people find the idea of the original position intuitively attractive when they hear of it for the first time. But I don't think this is because we all have a hidden talent for metaphysical inquiry. I think the idea of the original position hits the spot for the mundane reason that we recognize a device that we unconsciously use whenever appealing to a fairness norm in ordinary life.

Veil of ignorance. When a fairness judgment is to be made using the original position, Alice and Bob imagine themselves behind a *veil of ignorance* that conceals their identity. Behind the veil of ignorance, Alice thinks that she may turn out to be either Alice or Bob with equal probability, and the same goes for Bob. In this hypothetical state of ignorance, they imagine themselves bargaining about how to reach a compromise in whatever sharing problem they may face.

People think that an agreement reached in the original position will be fair because the distribution of advantage in whatever compromise Alice and Bob reach will seem determined as though by a lottery. Devil take the hindmost then becomes an unattractive principle, since you yourself might end up with the lottery ticket that assigns you to the rear.

How might the original position have evolved? Why should it be thought to implement the Golden Rule? How does the Golden Rule relate to our previous discussion of utilitarian and egalitarian norms? These questions shape the agenda for what comes next.

B. The Deep Structure of Fairness

Recognition of the Golden Rule seems to be universal in human societies. Is there any reason why evolution should have written such a principle into our genes? Some equilibrium selection devices are obviously necessary for social life to be possible, but why should something like the Golden Rule have evolved?

If the Golden Rule is understood as a simplified version of the device of the original position, I think an answer to this question can be found by asking why social animals evolved in the first place. This is generally thought to have been because food-sharing has survival value.

Sharing food. The vampire bats mentioned earlier provide an example. Unless a vampire bat can feed every sixty hours or so, it is likely to die. The advantages of sharing food among vampire bats are therefore strong—so strong that evolution has taught even unrelated bats to share blood on a reciprocal basis.

By sharing food, the bats are essentially *insuring* each other against hunger. Animals can't write insurance contracts in the human manner, and even if they could, they would have no legal system to which to appeal if one animal were to hold up on his or her contractual obligation to the other. But the reciprocity principle tells us that evolution can get round the problem of external enforcement if the animals interact together on a *repeated* basis.

By coordinating on a suitable equilibrium in their repeated game of life, two animals who are able to monitor each other's behavior sufficiently closely can achieve whatever could be achieved by negotiating a legally binding insurance contract. It will be easier for evolution to find its way to such an equilibrium if the animals are related, but the case of vampire bats shows that kinship isn't necessary if the evolutionary pressures are sufficiently strong.

Insuring against hunger. What considerations would Alice and Bob need to take into account when negotiating a similar mutual insurance pact?

Imagine a time before cooperative hunting had evolved, in which Alice and Bob foraged separately for food. Like vampire bats, they would sometimes come home lucky and sometimes unlucky. An insurance pact between them would specify how to share the available food on days when one was lucky and the other unlucky.

If Alice and Bob were rational players negotiating an insurance contract, they wouldn't know in advance who was going to be lucky and who unlucky on any given day on which the contract would be invoked. To keep things simple, suppose that both possibilities are equally likely. Alice and Bob can then be seen as bargaining behind a veil of uncertainty that conceals who is going to turn out to be Ms Lucky or Mr Unlucky. Both players then bargain on the assumption that they are as likely to end up holding the share assigned to Mr Unlucky as they are to end up holding the share assigned to Ms Lucky.

I think the obvious parallel between bargaining over such mutual insurance pacts and bargaining in the original position is no accident. To nail the similarity down completely, we need only give Alice and Bob new names when they take their places behind Rawls' veil of ignorance. Since we are offering a foundation myth in competition with the bible, Alice and Bob will be called Adam and Eve when they are behind the veil of ignorance.

Instead of Alice and Bob being uncertain about whether they will turn out to be Ms Lucky or Mr Unlucky, our story about the original position requires that Adam and Eve pretend to be ignorant about whether they will turn out to be Alice or Bob. It then becomes clear that a move to the device of the original position requires only that the players imagine themselves in the shoes of somebody else—either Alice or Bob—rather than in the shoes of one of their own possible future selves.

If Nature wired us up to solve the simple insurance problems that arise in food-sharing, she therefore also simultaneously provided much of the wiring necessary to operate the original position.

Of course, in an insurance contract, the parties to the agreement don't have to *pretend* that they might end up in somebody else's shoes. On the contrary, it is the reality of the prospect that they might turn out to be Ms Lucky or Mr Unlucky that motivates their writing a contract in the first place. But when the device of the original position is used to adjudicate fairness questions, then Adam knows perfectly well that he is actually Alice, and that it is physically impossible that he could become Bob. To use the device in the manner recommended by Rawls and Harsanyi, Adam therefore has to indulge in a counterfactual act of imagination. He can't become Bob, but he must pretend that he could. How is this gap between reality and pretense to be bridged without violating the Linnaean dictum that Nature doesn't make jumps?

Expanding Circles. As argued earlier, I think that human fairness norms arose from Nature's attempt to solve certain equilibrium selection problems. But Nature doesn't jump from the simple to the complex in a single bound. She tinkers with existing structures rather than creating hopeful monsters. To make a naturalistic origin for the device of the original position plausible, it is therefore necessary to give some account of what tinkering she might have done—just as evolutionary biologists explain the evolution of the eye by a sequence of little steps starting from light-sensitive spots on an animal's skin.

In Peter Singer's *Expanding Circle*, the circle that expands is the domain within which norms are understood to apply. For example, if you ask Swedes how they came to drive on the right like the rest of continental Europe, they will say it is because that has been the law in Sweden since 3 September 1967. But the law doesn't need much enforcing because driving on the right is an equilibrium in the Driving Game. A more controversial example uses the fact that it can still be in equilibrium for daughters who have ceased to love their mothers to care for them in their old age. The daughters will explain that they are fulfilling their moral duty by behaving as a loving daughter should, but would they be as dutiful if it weren't in equilibrium to be dutiful? Jesus sought similarly to expand the domain of the principle that you should love your neighbor by redefining a neighbor to be anyone at all, but he met with only limited success because it is only rarely that acting as though you love a stranger is an equilibrium strategy.

How does evolution expand the domain within which a moral rule operates? My guess is that the domain of a moral rule sometimes expands when players misread signals from their environment, and so unconsciously apply a piece of behavior or a way of thinking that has evolved for use within some inner circle to a larger set of people, or to a new game. When such a mistake is made, the players attempt to play their part in sustaining an equilibrium in the inner-circle game without fully appreciating that the outer-circle game has different rules. For example, Alice might treat Bob as a sibling even though they are unrelated. Or she might treat a one-shot game as though it were going to be repeated indefinitely often.

A strategy profile that is an equilibrium for an inner-circle game won't normally be an equilibrium for an outer-circle game. A rule that selects an equilibrium strategy in an inner-circle game will therefore normally be selected against if used in an outer-circle game. But there will be exceptions. When playing an outer-circle game as though it were an inner-circle game, the players will sometimes happen to coordinate on an equilibrium of the outer-circle game. The group will then have stumbled upon an equilibrium selection device for the outer-circle game. This device consists of the players behaving *as though* they were constrained by the rules of the inner-circle game, when the rules by which they are actually constrained are those of the outer-circle game.

Family life. Everybody agrees that the origins of human sociality are to be found within the family. A game theorist will offer the explanation that the equilibrium selection problem is easier for evolution to solve in such games. The reason why is to be found in Hamilton's rule, which explains that animals should be expected to care about a relative in proportion to their degree of relationship to the relative. Family relationships therefore provide a natural basis for making the kind of interpersonal comparison of utility that is necessary to operate the device of the original position.

The circle was then ready to be expanded by including strangers in the game by treating them as honorary or fictive kinfolk, starting with outsiders adopted into the clan by marriage or cooption. Indeed, if you only interact on a regular basis with kinfolk, what other template for behavior would be available?

Putting yourself in another's shoes. The next step is to ask how the original position gets to be used not just in situations in which Alice and Bob might turn out to be Ms Lucky or Mr Unlucky, but in which they proceed as though it were possible for each to occupy the role of the other. To accept that I may be unlucky may seem a long way from contemplating the possibility that I might become another person in another body, but is the difference really so great? After all, there is a sense in which none of us are the same person when comfortable and well fed as when tired and hungry. In different circumstances, we reveal different personalities and want different things.

To pursue this point, consider what is involved when rational players consider the various contingencies that may arise when planning ahead. To assess these, players compute their expected utility as a weighted average of the payoffs of all the future people—lucky or unlucky—that they might turn out to be after the dice has ceased to roll. When choosing a strategy in a family game, players similarly take their payoffs to be a weighted average of the fitnesses of everybody in their family. In order to convert our ability to negotiate insurance contracts into a capacity for using fairness as a more general coordinating device in the game of life, all that is then needed is for us to hybridize these two processes by allowing players to replace one of the future persons that a roll of the dice might reveal them to be, by a person in another body. The empathetic preferences that are needed to assess this possibility require nothing more than that they treat this person in another body in much the same way that they treat their sisters, cousins or aunts.

Deep structure? The preceding account of the possible evolution of the original position has a similar status to Chomsky's deep structure of language. That is to say, the original position models the deep structure of human fairness norms. So why do different societies have different views—not on the importance of fairness itself—but on the details of what counts as fair? How come, for example, that Aristotle thought slavery was fine as long as only barbarians were enslaved? Or that Spinoza thought that women should be treated as a lesser breed?

To find an answer, we need to ask where Alice and Bob get their empathetic preferences from in the original position. They can't do without them—otherwise they wouldn't be able to weigh up prospects in which they have to compare being Alice in one situation with being Bob in another. In fact, it is hard to think of a reason why evolution equipped us with the expensive mental capacity to feel empathetic preferences, if not to serve as inputs when making fairness judgments.

But the particular empathetic preferences Alice and Bob may have in a particular time and place are social constructs—like the actual languages spoken in different parts of the world. We learn to adopt one set of empathetic preferences rather than another by imitating the behavior of those we hope to emulate when we see them acting fairly. This is why Aristotle assigned a small social index to barbarians and Spinoza to women. They thought they were simply acting naturally, but they were actually the victims of the prejudices built into the social consensus of their respective societies—just as we are similarly the victims of the social prejudices built into our own social consensus today.

Analogy with language. The deep structure of language is universal in the human species and therefore presumably coded in our genes. But the languages spoken in different countries vary because the particular language spoken in a particular country is a product of social evolution—it depends on the cultural history of the people who speak it. The deep structure of language with which we are biologically programmed is necessary to allow babies to learn a language, but the particular language they learn depends on where they are brought up.

I think the same is true of fairness. We are biologically programmed with the deep structure of fairness, but the shape of the particular fairness norms we learn as we grow up is determined by the standard of interpersonal comparison that social evolution has crafted in the particular society in which we live.

David Hume said much the same thing about social norms in general. The Natural Laws to which writers of footnotes to Plato appeal are actually artificial in the sense that they are products of social evolution. What is natural is that we have such artificial laws. Similarly empathetic preferences are social constructs. What is determined biologically is that we are able to entertain such empathetic preferences.

III. Utilitarianism or Egalitarianism?

If Alice and Bob use the original position to make fairness judgments, will they act like utilitarians or egalitarians? Harsanyi says the former and Rawls the latter.

Both Harsanyi and Rawls assume the existence of a social consensus on how interpersonal comparisons are to be made. This assumption trivializes the bargaining between Adam and Eve in the original position. They both have the same empathetic preferences and so will simply agree on whatever they both like best. We therefore might as well replace them by Carol. The result will then be utilitarian.

So how come Rawls claims the result will be egalitarian? He gets to this conclusion by replacing expected utility by the maximin criterion (so that Adam and Eve seek to maximize the utility of whichever of Alice and Bob ends up with the smaller utility in their agreed outcome). Egalitarianism follows immediately, provided that the set of utility pairs from which a fair outcome is to be selected is convex.[2]

I guess Rawls didn't understand how iconoclastic it is to reject Bayesian decision theory. If it is to be rejected in the small world of the original position, it would need to be rejected across the board. One might as well reject arithmetic because you don't like what it says about how much tax you owe! The real difference between utilitarianism and egalitarianism is to be found in asking why people honor fairness norms at all.

Strains of commitment. Allegorical statues representing justice take the form of a blindfolded matron holding aloft a pair of scales and a sword. Her blindfold can be taken to correspond to the veil of ignorance. She needs her pair of scales to weigh Alice's worth against Bob's. She needs her sword to chastise Alice or Bob if they decline to accept her judgment about what is fair. As Thomas Hobbes put it: Covenants without the sword are but words.

This blunt observation is typical of the irrepressible Hobbes. Traditionalists don't like it at all. They think talk of power in a moral discussion is totally misplaced. Harsanyi and Rawls thought the same. Their answer to the question of why fairness norms should be honored is to invent yet more skyhooks like such traditional skyhooks as Practical Reason, Moral Intuition, General Will, or Natural Law. Harsanyi calls his skyhook Moral Commitment. Rawls calls his Natural Duty—but later questions his own invention by writing at length on the "strains of commitment" that such notions impose on Alice and Bob. I am with Hobbes on this question. Gurus may tell us that we have a Natural Duty to play fair, but why should we listen? If Natural Duty doesn't promote fitness, evolution will pay no attention at all.

It is enforcement that distinguishes when egalitarianism applies rather than utilitarianism. The latter applies when the sources of authority invented by philosophers are replaced by a *real-world* external enforcement agency that compels obedience to whatever a fairness norm specifies. This is why it made sense for Carol to be a utilitarian when asked to reform the tax code. She had the government standing by to police her new code. How many people would report their taxes honestly if we relied on a taxpayer's sense of natural duty for this purpose? Cheating is widespread even with the government doing its best to stop it.

[2] The reciprocity principle guarantees convexity provided that only equilibria of our repeated game of life are regarded as feasible.

Self-policing fairness. Why is egalitarianism appropriate when external enforcement is unavailable?

In talking about expanding circles, it was proposed in Part. II.B that we sometimes play real-life games as though bound by rules that don't actually apply. Such (largely unconscious) behavior can survive if it results in an equilibrium being chosen from those available in the game actually being played. However, the rules of the inner-circle game that survives as an equilibrium-selection device in the outer-circle game that is currently being played need to be realistic. We mustn't follow the tradition in philosophy of making up whatever rules confirm our prejudices. The inner-circle game must be a game that our ancestors would recognize as making sense in the environment in which they lived. In particular, it must operate in the absence of any external enforcement.

In examining the device of the original position from this point of view, there are three points at which external enforcement needs to be excluded:

1. The first point requires an appeal to the reciprocity principle of a previous chapter of my forthcoming book. The agreements envisaged as being reached in the original position must be viable as equilibrium outcomes in the ongoing game of life that Alice and Bob are actually playing. In particular, we must remember that property rights are a human invention. As David Hume explains, it is only by convention that you own your home, or your car, or anything else. John Locke tells us that we have a Natural Right to our own bodies, but as someone who invested money in the slave trade of his time, he should have known better.

2. The bargaining envisaged in the original position must be realistic. We mustn't cheat by inventing some fair process imported from the golden age of the poets. However, we can leave this issue until the next section, since our current assumption that Alice and Bob have the same empathetic preferences trivializes the bargaining problem.

3. It matters how we model the random event that determines which of Adam and Eve will turn out to be Alice and Bob when they supposedly emerge from behind the veil of ignorance. This very simple issue turns out to be the fulcrum on which a defence of egalitarianism turns.

The lot causeth contention to cease. The Book of Proverbs endorses the tossing of a coin to settle disputes. Adam and Eve do the same when considering who will be Alice and who will be Bob when they leave the original position. But the coin tossed in the original position is only hypothetical. When is this phantom coin tossed? How do they know which way it falls?[3]

[3] The actual probability with which these two alternatives occur isn't significant, because we can absorb any deviation from our ongoing assumption that the two alternatives are equally likely into Alice

In the following example, Alice and Bob both need a heart transplant, but only one heart is available. If the lives of both are regarded as equally valuable, then a utilitarian will be indifferent between giving the heart to Alice or Bob. But Alice would regard it as grossly unfair if the heart were then given to Bob on the grounds that he is a man—or because he is white or rich. Nor would she be at all mollified if told that she had an equal chance of being a man when the egg from which she grew was fertilized in her mother's womb. As a schoolboy, I was similarly made to sing the hymn that goes:

The rich man in his castle,
The poor man at his gate,
God made them high or lowly,
And ordered their estate.

The implication is that God tossed the coin that decided our social status, and so Alice has no just grounds for complaint in finding herself among the lowly.

Alice therefore cares a lot about *when* the phantom coin is tossed. If such a random event is to be used to determine who gets the heart, she will argue that a real coin should be tossed right now. If the coin falls in favor of Bob, she will be tempted to find arguments why this particular toss is unfair. So how is it possible to gain her unforced consent to honoring the fall of a coin that is only hypothetical?

The very simple answer is that for the original position to be viable without external enforcement, Adam and Eve must be indifferent between the coin falling heads or tails. Otherwise, if the original position were played for real, either Adam or Eve would reject the outcome and call for the coin to be tossed again.

Egalitarianism justified! When Adam and Eve have the same empathetic preferences—as we are currently assuming—the outcome of using the original position will therefore lead to the standard conception of an egalitarian sharing rule. Alice and Bob will get whatever share equalizes their empathetic utilities. When their shares are measured in terms of their personal utilities, these utilities will therefore be proportional to their social indices.

So we begin and end with Aristotle. He tells us first that the "sources and springs of friendship, political organization, and justice" lie within the family, and also that "What is fair . . . is what is proportional."

and Bob's social indices. Being a mighty hunter will therefore be one of the criteria that generate a larger social index.

When are people actually egalitarian? The preceding argument for fairness norms having an egalitarian structure applies particularly well to the ancestral bands of hunter-gatherers in which evolution shaped the human way of being social. Anthropologists are solidly of the view that the sharing of food—especially meat—in such groups was very egalitarian. But we don't know to what extent such sharing could be modeled as proportional to a person's social index. It was only late in the day that anthropologists thought to measure how much different people got when meat was shared, and the available data is inadequate to settle the issue.

There are also laboratory experiments carried out by psychologists under the heading of "modern equity theory". Their theory is the same as the egalitarianism of this essay, but derived independently.[4] The theory is said to be modern to distinguish it from Aristotle's observation that what is fair is what is proportional.

Do people honor egalitarian norms in the laboratory? The evidence is mixed. The theory does well in cases when Alice and Bob are said to have jointly invested in a business project and the time has now come to split the profits—which are then split in proportion to the size of each player's investment. The worst cases I know occur when Alice and Bob are friends or lovers. If psychologists are reading this, I hope they will consider new experiments on modern equity theory that pay closer attention to what counts as a reasonable correlate of utility in different contexts and what social indices may be in use.

Finally, I want to draw attention to some theoretical results from cooperative game theory in which axioms are proposed in an attempt to characterize what counts as fair. All the axiom systems of which I know that postulate full interpersonal comparison—of both the zero and the unit on Alice and Bob's utility scales—yield the same conclusion, variously called the egalitarian or proportional bargaining solution because it is identical to what we have been calling an egalitarian norm.

IV. Achieving a Social Consensus

Why has evolution equipped us with the capacity to have empathetic preferences? So that we can make fairness judgments. When do they take the form of a weighted sum of Alice and Bob's personal utilities? When both are completely successful in their attempt to empathize with the other. So why aren't our fairness norms utilitarian? Because utilitarian norms require external enforcement. Why are our fairness norms egalitarian instead? Because

[4] For broadly philosophical discussions, see Graham F. Wagstaff, An Integrated Psychological and Philosophical Approach to Justice: Equity and Desert (E. Mellen Press 2001); George Caspar Homans, Social Behavior: Its Elementary Forms (Harcourt, Brace & World 1961).

when a social consensus on empathetic preferences exists, the only agreement in the original position that doesn't somehow rely on external enforcement equalizes their empathetic utilities. Why do we need the original position at all? Because it still works when there is no social consensus, and so Alice and Bob may have different empathetic preferences. Why will social evolution remove differences between our empathetic preferences when the original position is in use? That is the subject of this section.

No secrets. There is no depth to the answers given to these questions except for the last—provided that we are willing to go along with the assumption that Alice and Bob have no secrets from each other. This is a strong assumption in a modern society and it is doubtless because it often fails that we aren't more successful in using fairness to solve disputes. However, we shall continue to maintain the assumption in this section because without it we wouldn't be able to appeal to the following reasonably realistic model of bargaining when trying to work out what deal Adam and Eve will reach behind the veil of ignorance when they have different empathetic preferences. But we first need to be more careful about what kind of sharing problem Alice and Bob need a fairness norm to solve.

Sharing Game. A kind of sharing game was used as an example when comparing egalitarianism and utilitarianism in a previous chapter of my forthcoming book. We looked at the very different results of using these two conceptions of how fairness norms work in the case when Carol is asked to share a sack of flour between Alice and Bob. Essentially the same model of sharing is recycled here, but without Carol to provide a common standard of interpersonal comparison.

Anthropologists report that meat obtained by cooperative hunting in hunter-gatherer groups is commonly shared very fairly. But the problem of assessing the costs and benefits of such cooperative behavior will be abstracted away. When the benefits don't outweigh the costs, cooperative behavior won't evolve.

With this caveat, we might as well proceed as though Alice and Bob have come across a dead animal, and the question is how they are to share the meat. What would happen if they were unable to agree on how to share then matters. We have to deal with the problem envisaged by Epicurus when he said that fairness exists to coordinate behavior on some compromise as an alternative to getting into a fight. It is necessary to specify how Alice and Bob evaluate the (possibly violent) consequences of a failure to agree.

We can summarize these considerations by saying that they reduce the sharing question to what economists call a bargaining problem. This simply consists of a set of possible agreements and a disagreement outcome—often called the status quo—from which Alice and Bob can depart only by mutual consent. Both the possible agreements and the disagreement outcome are represented by the expected utilities assigned to them by Alice and Bob.

Bargaining. The bargaining envisaged as taking place behind the veil of ignorance must be the ordinary kind of bargaining to which Alice and Bob are accustomed in real life.[5] How else could they predict what its outcome would be? So we don't want any fancy bells or whistles. We especially don't want any suggestion that only fair bargaining is allowed—otherwise our attempt to explain how fairness norms work would be circular. Plato's Thrasymachus would therefore have approved of the bargaining theory needed: the bargainers negotiate face-to-face, bringing to bear whatever power they may have at their disposal.[6]

The economic theory of bargaining invented by John Nash and Ariel Rubinstein for the case when the bargainers have no secrets from each other is therefore just what is required. To cut what could be a long story short, both their approaches lead to what is called the Nash bargaining solution.[7]

Rubinstein's bargaining model is the more realistic. In a simple version, Alice and Bob alternate in making offers on how to split whatever is available. The bargaining ends either when an offer is accepted, or some small probability event outside anyone's control interrupts the bargaining, in which case Alice and Bob are stuck with the disagreement outcome. This bargaining game has a unique perfect equilibrium that approximates the Nash bargaining solution when the interruption probability after each refusal of an offer gets small.

[5] I am making a fuss about this obvious point because the nearest orthodox philosophy that comes to the approach described here is David Gauthier, Morals by Agreement (Oxford Univ. Press 1986), in which he not only argues that cooperation is rational in the Prisoners' Dilemma, but invents his own theory of bargaining.

[6] In Plato's *Republic*, Glaucon gets short shrift when he proposes something like the reciprocity principle as the mechanism that holds societies together. On the other hand, Plato's fictional Socrates takes a lot of time over the easy task of refuting Thrasymachus when he absurdly suggests that talk of justice simply disguises the unbridled exercise of power.

[7] John Nash is famous for his 1951 definition of a Nash equilibrium, although he was anticipated by Auguste Cournot in 1838. But Nash was truly original in his 1950 thoughts on bargaining, which economists at the time dismissed as a neglected branch of psychology. He gave a list of rationality assumptions that determine a unique outcome of a bargaining problem called the Nash bargaining solution—which is not at all the same thing as a Nash equilibrium. He supported this result by analyzing a simple bargaining model called the Nash demand game, in which Alice and Bob make simultaneous demands, which are implemented if jointly feasible and which result in the disagreement outcome otherwise. With a little uncertainty about what potential agreements are actually feasible, he found that all Nash equilibria of his demand game approximate the Nash bargaining solution. Economists largely ignored this discovery until 1982, when Ariel Rubinstein reported that his much more realistic alternating-offers bargaining game has a unique perfect equilibrium. My own contribution was to show that this unique bargaining outcome implements a generalized version of the Nash bargaining solution. All this is explained in my book *Playing for Real* with a minimum of mathematics, and in my *Natural Justice* with some diagrams substituting for equations. Since we are not using mathematics at all here, much of what I shall be saying will have to be taken on trust.

A simple example may help give the flavor of how the Nash bargaining solution works. Suppose that Alice and Bob are a divorcing couple in dispute over who gets custody of their baby without the opportunity to appeal to a court of law. If they can't agree, the baby goes into an orphanage, which each regard as the same as the other getting the baby. Various compromises are available. They could agree to some time-sharing arrangement, but it is simpler to assume that Alice and Bob only consider agreements that say who gets the baby with what probability. The Nash bargaining solution unremarkably says that each will then get the baby with probability one half.

For this reason, some authors assess the Nash bargaining solution as if it were an attempt to characterize a fair arbitration scheme, but this is to miss the point altogether. It is totally without virtue as a fairness criterion because it remains the same however we compare Alice and Bob's personal utility scales. This is easy to see in the preceding example. Alice may love the baby very dearly and Bob hardly at all. If so, it wouldn't be fair that Bob should have an equal chance of getting the baby—which is why King Solomon had to find a way of determining who loved the baby best in his famous judgment.

Why play fair when you could bargain? In modern times, people use fairness norms when the number of players is too large or communication is too difficult for bargaining to be practical, or else because the cost of bargaining in time and trouble is too high. Sometimes an attempt at bargaining would be punished by the disapproval of onlookers—as for example at a dinner party when a dish in short supply is being shared. Fairness is often mentioned in questions of public policy, but this is mostly just for rhetorical purposes.

Things were more interesting in ancestral times. It seems likely that fairness is as ancient as language—perhaps more ancient. If so, then bargaining wasn't available as an alternative coordinating device. The other obvious alternative to fairness is leadership, but the hunter-gatherer societies that survived into modern times had no leaders. In fact, the punishments inflicted on tough guys who sought to dominate their fellows were sufficiently severe that anthropologists report that everybody understood that it was wise not even to give the impression of being bossy. Fairness was therefore much more important in human ancestral societies than it is to us—which goes a long way to explaining why their egalitarianism came as a surprise to early anthropologists.

Modeling social evolution. It is time to outline how social evolution might generate a social consensus on how interpersonal comparisons are made. Nothing profound is involved. People mostly have the same empathetic preferences for the same reason that teenagers who hang out together dress similarly and like the same kind of music—or that traditional philosophers all admire Plato and Kant—because the way to get ahead is to copy those who seem to have got ahead already. The following outline of a model fleshes out this story, but nothing much will be lost by skipping forward to Part V.

Memes. Richard Dawkins introduced the word *meme* to serve as a substitute for *gene* when biological evolution is replaced by social evolution. I don't think it helpful to pursue the biological parallel too closely, so my use of the term is better understood in the sense in which it has entered popular culture.

I treat empathetic preferences as memes propagated by imitation. Alice and Bob unconsciously alter their empathetic preferences if they would do better in terms of their personal utilities when the original position is used by copying the empathetic preferences of others. The players' strategies in the underlying evolutionary game then correspond to their (largely involuntary) choices of different standards of interpersonal comparison. At a Nash equilibrium of this evolutionary game, neither Alice nor Bob will have an incentive to switch from their current standard of interpersonal comparison to one they see being operated by others.[8]

Social consensus achieved! It turns out that the evolutionary process just described results in Alice and Bob ending up with the same empathetic preferences, and so they will agree on what social indices should be assigned to different types of people. We can therefore recycle our earlier analysis in which the existence of a social consensus was taken for granted. In particular, we should expect to see egalitarian fairness norms in operation in situations where no external enforcement agency is available.

Evolution erodes the moral content of norms? An equilibrium in empathetic preferences encapsulates the cultural history of a society that led people to adopt one standard of interpersonal comparison rather than another in a particular context. Traditionalists don't like the fact that this history will largely be shaped by the way in which power is distributed in whatever society is under study. In their fairy stories, power can't be relevant to how moral norms are shaped—morality is a substitute for power! Their worst fears are apparently confirmed when they learn the conclusion to which our current evolutionary model leads.

If the underlying sharing game Alice and Bob play in real life was always the same, social evolution would shape their empathetic preferences so that the egalitarian norm coincides with the Nash bargaining solution of the

[8] My book *Natural Justice* explains why different ways of modeling the imitation process can lead to different definitions of an empathy equilibrium. The more refined definition considered there requires that neither Adam nor Eve in the original position have an incentive to misrepresent their empathetic preferences when these are evaluated using the empathetic preferences that Alice and Bob actually hold. The refined definition leads to the same conclusion as the cruder version of the text.

sharing game.[9] But the Nash bargaining solution has no virtue as a fairness norm.

Does this mean that the traditionalists are right to argue that evolution and morality are like oil and water in that they cannot be mixed? It certainly does mean that we can't expect the kind of morality that evolved for use in real-life situations to coincide with traditional fantasies, but it doesn't imply that we can throw away fairness norms in favor of implicit bargaining models. The reason is that we use the *same* fairness norm to solve many *different* sharing games in real life.

In our simple model of social evolution, Alice and Bob's empathetic preferences will evolve so that the egalitarian norm of the historically *average* sharing game coincides with the Nash bargaining solution. In other sharing games—especially those that are unusually asymmetric or that incorporate relatively new technologies—the egalitarian outcome won't look like the Nash bargaining solution at all.

Traditionalists who think that empathy matters at all may therefore continue to complain that they don't like the way empathetic preferences are shaped by social evolution, but they can't simultaneously complain that they have no role in the fairness norms we use in real life. The needy will get more in contexts where hunger is an issue. The rich will pay higher taxes. And so on.[10] If such conclusions didn't follow, we wouldn't have come anywhere near explaining how evolution shaped our intuitions of how fairness works.

For those with a scientific background, the model has the additional merit that it provides a theoretical way of predicting what standards of interpersonal comparison different evolutionary histories will generate. All we need to do in principle is to work out the Nash bargaining solution of the historically average sharing game, and then choose Alice and Bob's social indices so that this coincides with the egalitarian outcome. As with much else, easier said than done!

Local justice. As Epicurus explained, justice is the same for all at any particular time and place, but may vary between different times and places—a fact that has been widely documented.[11] But one doesn't need to look very far to see

[9] In which case, both norms also coincide with the utilitarian norm. So social evolution will lead both egalitarians and utilitarians to the same standard of interpersonal comparison. Perhaps this is why Harsanyi and Rawls favored similar reforms in private conversation, although espousing very different ideas on how societies should be organized.

[10] My *Natural Justice* pursues the impact of changes in need, effort, ability and status on a person's social index by looking at how altering suitably defined versions of these notions changes the historical average sharing game.

[11] *E.g.*, Jon Elster, Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens (Russell Sage Found. 1992); H. Peyton Young, Equity: In Theory and Practice (Princeton Univ. Press 1994). It is usually taken for granted that the standard of interpersonal comparison is universal, and so the differences are attributed to the method used to compute a fair outcome. In the approach

that we take this for granted when making fairness judgments. For example, everybody agrees that need is what matters when assigning food stamps, but merit is what counts when awarding Nobel prizes.

In our evolutionary model, this phenomenon is captured by asking who Alice and Bob copy when adjusting their empathetic preferences. If they only copy people they see operating the original position in a particular context, then the standard of interpersonal comparison that evolves in that context needn't resemble the standard that evolves in other contexts.

V. Social Contracts

I hope the discussion so far has at least made it plausible that evolution could have generated the fairness norms that actually get used in real life for solving small-scale coordination problems. Just as it would have been better if evolution had made us stronger and cleverer, so it would perhaps have been better if evolution had made our social life more like that of bees and ants, but we are stuck with what evolution actually made of us.

But we aren't altogether helpless victims of our evolutionary history. Our big brains evolved partly so that we could chip flint to make stone tools, but we used these tool-making skills to create modern technology. Our big brains also evolved to allow us to live amicably together in small societies without anyone ordering us around. Perhaps we can similarly use these social skills to learn to live together more amicably in large societies.

Scaling up? Can we learn to use the fairness norms that evolved for small-scale applications to solve large-scale problems? In doing so, we must put aside the grandiose aspirations of writers of footnotes to Plato. They have been telling us to ignore what evolution made of us for more than two thousand years without any notable success. The fairness norms we actually use in real life aren't shadows cast by some absolute noumenal world; they are simply social tools washed up on our evolutionary beach. If enough of us sufficiently near the levers of power want to use them to improve how big societies work, let us just get on and do it. We don't need to invent the kind of metaphysical justifications that traditionalists think necessary. It is enough that we want to do it. Perhaps we would want to do something else if our histories—both personal and social—had been different, but so what?

defended here, the original position is assumed to be universal, and the differences attributed to variations in the standard of interpersonal comparison.

A. Traditional Social Contracts

Before exploring the possibility of using egalitarian fairness norms to improve our current social contract, it may be helpful to look at some of the traditional approaches to thinking about social contracts. Two aims need to be distinguished. Epicurus would have approved of the first. It seeks to explain how societies came to be as they are without inventing metaphysical skyhooks. He would have been less enthusiastic about the second aim, which is to justify replacing our current social contract with some utopian alternative. He would have been even less enthusiastic about how the utopian aim is often muddled with the historical aim.

Social contracts as historical constructs. The original philosophical approach was to suggest that the rules which govern our moral and political behavior were agreed at some ancient conclave, whose authority we are still somehow obliged to honor. David Hume wrote an essay condemning this notion of an "original contract". Why are we be bound by agreements made at such a conclave? Where is the evidence for such an ancient conclave anyway?

Hume accepts that talk of an ancient conclave may stand in for a history of lesser coordinating innovations that gained general acquiescence over time, but points out that the histories of actual constitutions are largely accounts of usurpation, conquest and rebellion. He has no time at all for the fiction that we could somehow be bound by some historical agreement of whose very existence we have no knowledge.

How are Hume's criticisms of the idea of the original contract to be reconciled with his much admired views on reciprocity and the evolution of conventions? We must remember that Hume didn't have modern game theory at his disposal, and so didn't realize the extent to which the reciprocity principle can be used to sustain cooperative equilibria in repeated games. When operating such cooperative equilibria, we don't need to invent metaphysical reasons to justify our behaving cooperatively because, most of the time, it is in our long-run self-interest to do so.

Nor did Hume know of Darwin's theory of evolution. He says somewhere that species never change. So when he speaks of history, he isn't referring to evolutionary history. If we ignore ephemeral aspects of social contracts, like which family counts as royal or which side of the road on which to drive, his analysis is therefore unproblematic for the scientific approach to social contracts pursued in my forthcoming book—which I hope goes some way toward fulfilling the aspirations of the early social contract theorists like Grotius and Pufendorf whose ideas Hume was criticizing.

Social contract theory as a political tool. The idea of a social contract was a hot topic for David Hume because of its use by the Whigs of his day in opposing the Tories. The Tories were conservatives who favored retaining the traditional British monarchy because they thought this is how things had always been. The Whigs argued that governments operate by mutual consent of the governed, and so kings can legitimately be replaced if this consent is lost.[12] The Whigs were inspired by the social contract ideas of John Locke. The Tories didn't feel the need for any philosophy, but they could have appealed to Thomas Hobbes, whose earlier social contract theory defended the authority of kings.

Looking back, it is plain that social contract theorists like Hobbes and Locke began with their favored conclusion and invented whatever philosophical arguments took them to where they wanted to go—rather like Plato when he tacked on as an afterthought the idea that his idealized Spartan constitution could be deduced from his ramshackle definition of justice. In the disputation that followed, the explanatory aspirations of the earliest social contract theorists were lost sight of altogether.

Hobbes' social contract. Thomas Hobbes (1588–1679) was a royalist who thought the rebels in the English Civil War irrational to prefer the brutality of the war to submission to the king. His *Leviathan* idealizes this political position by comparing the safety to be found in accepting the role of a cog in an authoritarian social machine with what he called the *state of nature*, which he famously characterized as a "war of all against all" in which life is "poor, nasty, solitary, brutish, and short". Everybody would doubtless agree if this were indeed the choice!

However, the immediate point is that the imaginary history of the human race that Hobbes invents is just a fable, but crooked thinkers often contrive to treat it as though it were intended to be taken literally while offering the faintest of praise for Hobbes' magnificent *Leviathan*.

Locke's social contract. John Locke (1632–1704) was even more of an empiricist than Hobbes. His idea that the human mind is a blank slate on which experience can write anything whatever continues to be popular with utopians in spite of all the objective evidence to the contrary from biology and neuroscience. His contribution to social contract theory can be seen as a philosophical justification for replacing the outdated morality of medieval Christianity

[12] The Whigs are traditionally associated with the Glorious Revolution of 1688, in which the Catholic and authoritarian James II was expelled in favor of the Protestant and constitutionally minded William III. American history also boasted a Whig party, broadly similar in character to its British counterpart. It was vocal in its opposition to Andrew Jackson's authoritarian innovations in the use of the presidential veto. Before joining the newly emergent Republican party, Abraham Lincoln was a Whig, but modern Republicans have largely forgotten their whiggish roots.

by a morality more suited to the new commercial age in which he lived. In politics, these ideas became the bedrock on which British opposition to the unbridled authority of kings was based.

His fable of the origins of the social contract is perhaps best seen as a story of how a liberal constitution might have arisen in an ideal world. He introduces the skyhook of Natural Rights[13] according to which "No one ought to harm another in his life, health, liberty, or possessions". Property rights can then only be transferred in a civilized way. They are originally acquired by mixing one's labor with property in a state of nature that precedes ownership, provided "enough and as good is left in common for others".[14]

Rousseau's social contract. Bertrand Russell is doubtful whether Jean-Jacques Rousseau (1712–1778) can be considered a philosopher at all. But whatever his genuine merits as a philosopher, his writings have arguably been more influential than the writings of the rest of the philosophy profession put together. He was for the Jacobins of the French Revolution what John Locke was for the earlier Whigs across the English Channel, but written very much larger. Perhaps equally important, he marked the moment in which a wave of irrational romanticism began to displace the values of the enlightenment.

Rousseau's arguments in support of his social contract theory seem to me largely cosmetic. In his romantic state of nature, we were noble savages wandering at peace with the world in the primeval forest, but although we were "born free", we are now "everywhere in chains". His solution to our decline from his blissful state of nature is to persuade us to honor the General Will. To articulate the General Will, we need someone with sublime wisdom: otherwise we might get stuck with the Will of All, which is not at all the same thing. On the contrary, the wills of all citizens must be brought into accord with the will of this mythical sage.[15] I guess this is what Robespierre thought he was doing when presiding over the Terror after the French Revolution.

Immanuel Kant absurdly idolized Rousseau as the "Newton of the moral world", but I hope I will be excused from discussing what Kant's categorical imperative supposedly implies about social contracts.

[13] As for the Rights of Man, the tyrannical Maximilien Robespierre tells us that "Any law that violates the inalienable rights of man is essentially unjust and tyrannical; it is not a law at all." It is inadequate to say that such talk of inalienable or imprescriptible Natural Rights in documents like the 1789 French Declaration of the Rights of Man is nonsense: to quote Jeremy Bentham, it is "nonsense upon stilts".

[14] John Locke, Two Treatises of Government (1689). Written in opposition to John Rawls' *Theory of Justice*, Robert Nozick's *Anarchy, State and Utopia* uses Locke's defence of a liberal constitution to defend a decidedly illiberal form of libertarianism.

[15] Jean-Jacques Rousseau, Du Contrat Social [The Social Contract] (1762). David Hume selflessly helped Rousseau when this seriously neurotic show-off was banished from France, but his reward was to be accused of plotting against him. Rousseau's *Confessions* tells us that he abandoned his five babies one by one on the doorstep of an orphanage. Perhaps this is why his book *Emile* is so hilariously unrealistic about the realities of bringing up a child to be a good citizen.

Rawls' social contract. John Rawls tells us that his theory of justice is in the tradition of Locke, Rousseau and Kant. In the rest of this essay, I argue that Rawls' moral intuitions are better seen as idealizations to whole societies of the fairness norms that we observe being used to solve small-scale coordination problems in our everyday lives, and to which the early part of this essay was devoted. I think that all moral philosophers really get their intuitions from the same source, but Rawls is one of the few who isn't led badly astray by their belief that they have a hotline to some metaphysical world available only to those gifted with sublime wisdom.

B. Widening the Scope of Fairness Norms

The early part of this essay offers a putative evolutionary history of the fair social contracts of the small societies of hunter-gatherers into which the human species was divided ten thousand years or more ago. My contention is that we still use the same fairness norms in modern times to solve small-scale coordination problems.

I now move on from this historical approach to add my voice to the rhetorical efforts of social contract theorists like Hobbes and Locke. I propose seeking to use the egalitarian fairness norms that I think we use to solve small-scale problems to address the large-scale problem of improving our social contract. The idea will have political appeal only to the extent that it accords with the fairness intuitions that ordinary people acquire as they live their ordinary lives. In particular, the standard of interpersonal comparison employed must be whatever standard people actually use in real life, rather than some utopian ideal preached from a metaphysical pulpit.

Health example. The demand for health services is always going to exceed the supply. It is currently rationed in the United States largely by price. But how is rationing to be organized when the state provides health care supposedly for free? In Britain, it is customary to deny that there is any rationing at all while actually rationing health care by making people wait for treatment.

My own view is that people will find it acceptable to be told the truth about rationing—provided that the rationing is done in a way that they perceive as fair. But we won't be able to persuade the public that a rationing scheme is fair unless we first spend time and money in finding out what they regard as fair. If I am right about how fairness norms work, this will require empirical research into how they make interpersonal comparisons of utility—how they assign social indices to different types of people when making small-scale fairness judgments.

Does history matter? Neither Harsanyi nor Rawls saw any need to propose a state of nature in defending their utopian aspirations. They proceed as though history doesn't matter.

Thomas Paine thought the same in his fiercely contested debate with Edmund Burke after the French Revolution.[16] Paine argued that we should throw away our old social contract—root and branch—and design a new social contract that owes nothing whatever to the past. Burke argued that a social contract is much more than Robespierre and his fellow revolutionaries understood. You can write words on a piece of paper and call it a constitution, but you waste your time if the citizens of your society aren't ready to accept your piece of paper as a coordinating device. What determines how much the citizens are willing to accept? The attitudes they inherit from the past.

Ignoring history seems to me not just unwise but on the edge of incoherence—rather like the Irish peasant who was asked the way to Dublin and replied that if he were going to Dublin he wouldn't start from here. We have no choice but to start from where we are now. Insofar as my approach to social contract issues can be said to have a state of nature, it is therefore the social contract we are operating right now in the society in which we currently live. In the debate between Thomas Paine and Edmund Burke, I am therefore firmly on the side of Burke. The real question is: How can we improve what we have got already? What gains are possible relative to the current status quo?

Efficiency. Fairness isn't the only thing that matters when writing idealized models of social contracts. The first priority is that a social contract be sufficiently stable to survive for long enough to be recognized as a social contract. This requirement is met by insisting that only equilibria of the game of life played by a society are viable as possible social contracts.

The second priority is that a social contract be efficient, which means that nothing is wasted—no reform is possible in which everybody can gain. An anecdote about a one-lane bridge in Ithaca, New York may help to explain why efficiency is relevant. Professors from nearby Cornell University liked to tell how cars behaved very fairly at this bridge by taking turns to cross, shrugging off the observations from foreign professors that cars on the one-lane bridges with which they were familiar did the efficient thing by waiting until a gap appeared in the flow of oncoming cars and then crossing until a gap appeared in their own flow. However, the last time I heard a Cornell professor on this subject, he reported that the current norm was for four cars to

[16] Edmund Burke, Reflections of the Revolution in France (1790); Edmund Burke, An Appeal from the New to the Old Whigs (1791); Thomas Paine, Rights of Man (1791); Thomas Paine, Common Sense (1776). Paine was influential in provoking the War of American Independence, but the authors of the American Constitution were wise not to discard what they thought worth keeping.

go one way before four cars went the other. In the face of increasing congestion, social evolution was on the way to generating an efficient solution to the Ithaca bridge-crossing problem!

Efficient social contracts. A fable may help explain how evolution can generate efficient social contracts in general. Suppose that many identical small societies are operating one of two social contracts, *busy* and *idle*. If *busy* makes each member of a society that operates it fitter than the corresponding member of a society that operates *idle*, then here is an argument which says that *busy* will eventually come to predominate.

To say that a citizen is fitter in this context means that the citizen has a larger number of children on average. Societies operating social contract *busy* will therefore grow faster. Assuming societies cope with population growth by splitting off colonies which inherit the social contract of the parent society, we will then eventually observe large numbers of copies of societies operating social contract *busy* compared with those operating contract *idle*.[17]

Fairness comes after efficiency. Fairness norms, and other norms like ownership norms or sexual taboos, only arise at the third level of priority in this idealized approach. They exist as selection devices when multiple efficient equilibria are available—which is all of the time when repeated games are involved.

Utopian reform? It may be helpful to summarize in three steps what would be involved in consciously attempting a fair reform of some aspect of a society. The emphasis here is on the difficulties along the way, both theoretical and practical.

1. The first step is to identify what reforms are feasible—which reforms lead to outcomes that won't fall apart because they don't respect the nature of human nature. Saintly folk like St Francis of Assisi are few and far between. A reform that requires people to neglect their own self-interest may work for a while, but it will gradually unravel as people invent reasons why it is OK to cheat a little here and there. In our idealizations, this stability requirement is met by requiring that only equilibria of the human game of life be considered. It must then be remembered that the politicians who run our governments and the officials charged with implementing their decisions are also players in the game of life. When can they be trusted? Only when guardians are in

[17] Although selection takes place among groups, the argument isn't an example of the group selection fallacy, because a social contract is identified with an *equilibrium* of the game of life played by each of the competing societies. But selection among equilibria doesn't require that individuals sacrifice anything for the public good, because every individual in every group is already optimizing his or her fitness by acting in accordance with the social contract of their society. The paradigm of the selfish gene is therefore maintained throughout.

place—along with guardians to guard the guardians. In brief, things need to organize so that the guardians guard each other.

It is tempting to respond that even this first step in implementing a fair reform is an impossible task, and it is true that the best one can hope for in real life is some temporary stability. Who knows where social evolution will eventually take a society after a reform has been implemented? But one can at least make the effort to look ahead and attempt to predict how a new system might be gamed. The alternative—which we usually see in practice—is equivalent to strapping on a pair of ramshackle wings and jumping off a high building in an attempt to fly.

2. The next step is to question the efficiency of possible reforms that pass the stability test. Is there some way we could improve on the proposed reform so that everybody gains? It is important here to recognize that creativity and enterprise mustn't be stifled as in the old Soviet Union, which collapsed largely because its economy was strangled by hopeless attempts to plan everything in advance. We are going the same way too, with rules and regulations coming out of our ears. In brief, decisions need to be delegated to the level at which the necessary knowledge and expertise resides.

There is also a major philosophical question: When it is said that efficiency requires that everybody should gain, who counts as everybody? What of mutual consent in general? Whose consent are we talking about? I have heard it argued that chimpanzees should count. If so, why not animals in general? What of young children, or people suffering from dementia or some other disabling mental illness? What of the economically powerless—the huddled masses of which Marx spoke so eloquently? What of people of another religion or a different color?

Here is yet another problem that writers of footnotes to Plato are reluctant to confront. The unwelcome truth is that only those who are able to influence the outcome actually count. If a group that is currently discriminated against can't organize itself to bring its collective power to bear, then it will have to rely on the sympathetic preferences of those who can. History is not very encouraging on the latter possibility. Recall that Aristotle thought that barbarians are natural slaves, and Spinoza that women lack the capacity for serious thought. Today the super-rich think the same about inferior beings like us who actually pay taxes.

A proper analysis of such questions will have to await a better understanding of both human psychology and the game theory of coalition formation. In the interim, we have to live with the unpleasant fact that mutual consent in practice actually means the consent of those sufficiently powerful—either individually or collectively—to influence what reforms have a chance of getting implemented.

Finally, there is the problem posed by the fact that people often don't even know who their neighbors are in a large modern society. The reciprocity principle says that a multitude of efficient equilibria is always available in an indefinitely repeated game—provided that everybody's business is an open book. But what happens when people have secrets from each other? The best theory has to offer at present are examples in which the extent to which reciprocity works depends on how much the players in a game know about each other. However, the currently unregulated invasion of our privacy through the internet may paradoxically ease this difficulty (although I am no more keen on my own privacy being invaded than anybody else).

In any case, even when the kind of first-best efficiency available when information is not a problem is beyond our reach, we still need to strive for whatever second-best kind of efficiency can be achieved when information is a problem. Much to the distress of some writers of footnotes to Plato, markets will be part of the solution to this efficiency problem.[18] When working well, they can be marvellously effective in overcoming informational difficulties.

However, it is necessary to be skeptical about the claims of right-wing economists that markets are always the solution to everything. The theoretical results on which they rely are subject to all kinds of restrictions—there must be a large number of small buyers and sellers, goods must be divisible, preferences must be convex, and so on. When such restrictions are not satisfied—which is much of the time—markets need to be regulated. It is true, as right-wingers complain, that current regulation is often appallingly bad. But it would be a lot less bad if it weren't left in the hands of lawyers, who are often entirely ignorant of the most elementary economic principles.[19] Sometimes regulation is entrusted to the very capitalists whose activities are supposedly being regulated. It is very frustrating for game theorists when they know how a particular industry needs to be regulated, but their expertise on such social engineering is called upon only when lawyers want to discredit the arguments of other lawyers.

3. What remains to be done is to apply these insights to reform large-scale policies in the real world. We face no shortage of opportunities to do better—from global warming and health care to the setting of "fair, reasonable, and nondiscriminatory" (FRAND) royalties for standard-essential patents for

[18] Here is Robert Burton on markets: What's the market? a place (according to Anacharsis) wherein they cozen one another, a trap: nay, what's the world it self? A vast *chaos*, a confusion of manners, as fickle as the air, *domicilium insanorum*, a turbulent troupe full of impurities. a mart of walking spirits, goblins, the theatre of of hypocrosie, a shop of knavery, a nursery of villainy, the scene of babling, the school of giddiness, the academy of vice—and much more of the same!

[19] I was once employed as a consultant by a leading lawyer in economic regulation, to whom I was trying to explain that the short-haul UK package-holiday business might usefully be modeled as Bertrand-Edgeworth competition. To explain the idea, I began by saying that the idea is a hybrid of the Cournot and Bertrand models of oligopoly. He then asked: what is an oligopoly?

smartphones and other wireless devices.[20] But this task of applying fairness norms I must leave to my forthcoming book—and ultimately its readers.

Conclusion

In this essay from my forthcoming book, *Crooked Thinking or Straight Talk? Epicurus Shows the Way*, I have discussed evolutionary ethics, the Golden Rule, utilitarianism and egalitarianism, social consensus, traditional social contracts, and the widening scope of fairness norms. I have proposed seeking to use the egalitarian fairness norms that I think we use to solve small-scale problems to address the large-scale problem of improving our social contract. I have explained that fairness comes after efficiency—that is, fairness norms only arise at the third level of priority in the idealized approach that I describe. They exist as selection devices when multiple equilibria are available—which is all of the time when repeated games are involved. The real questions to be answered are: How can we improve what we have got already? What gains are possible relative to the current status quo?

[20] This opportunity to apply fairness norms to intellectual property recently came to my attention upon reading J. Gregory Sidak, *What Makes FRAND Fair? The Just Price, Contract Formation, and the Division of Surplus from Voluntary Exchange*, 4 Criterion J. on Innovation 701 (2019).