

Chapter 3

Two Problems of Cooperation

Katharine Browne

3.1 Introduction

Nature is rife with examples of cooperative behaviours. From honeybee societies and altruistic vampire bats to the enormously complex civilizations formed by human beings, cooperation lies at the foundation of all social interactions.

Finding examples of cooperation is not difficult. Explaining how it is possible, however, is less straightforward. The prevalence of cooperation in nature is a bit of a mystery, in light of evolutionary theory. The reason is that cooperation (as it is sometimes characterized) involves a fitness cost to its actor, a characteristic that is difficult to make sense of given that nature selects against fitness-decreasing traits. There is an analogous problem in normative moral theory. Cooperative behaviour (as it will be understood here) requires that individuals constrain self-interested pursuits. But ‘rationality’ (as it is most commonly understood) requires that agents act as selfish utility-maximizers. It thus seems that cooperative behaviour is at odds with rational behaviour.

This paper has two central aims. The first is to outline these two contextually different problems of cooperation, and provide solutions to each. My primary concern in the biological context will be with describing the possibility of *human* cooperation. What is distinctive about human cooperation is its scale and scope. And these characteristics make it particularly difficult to reconcile with the usual evolutionary explanatory mechanisms. Given that human cooperation extends beyond kin and occurs on a very large scale, the descriptive account will explore what kinds of mechanisms might support cooperation, and whether there is an evolutionary story that can accommodate these. I will defend an explanatory account of the emergence of cooperative behaviours that appeals to cultural group selection.

K. Browne (✉)

Centre for the Study of Mind in Nature (CSMN), University of Oslo, Oslo, Norway
e-mail: katharine.browne@csmn.uio.no

In the normative context, I will defend the rather unpopular view of the rationality of adopting the ‘cooperative’ disposition to constrained maximization, as articulated by David Gauthier. This defence requires two moves. The first is to show that the disposition to constrained maximization will yield a greater utility than an alternative disposition. The second is to make the case that the rationality of a disposition entails the rationality of the actions that the disposition recommends. I argue that Gauthier succeeds in both. And if so, we will have a reconciliation between cooperation and rationality.

The second aim of this paper will be to outline how the descriptive and normative projects are connected. One way in which the two problems of cooperation are related is through the social contract tradition. Social contract theory traces moral and political obligations to a contract. When individuals need to band together in cooperative ways, they agree to a set of principles that will govern their social interactions. These principles set the terms for cooperation. Within the contract tradition there is a descriptive branch and a normative branch. Descriptive approaches describe the origin of the social contract and seek to explain how cooperation occurs. The first problem of cooperation identified above falls to this branch. Normative approaches to the contract, by contrast, aim to justify the terms of the contract. Our second problem of cooperation falls to this branch. I will argue that there is a convergence between the descriptive and normative strands, specifically between what evolution produces and reason recommends. I will show that the cultural group selection explanation of the emergence of cooperation provides an explanation of the emergence of dispositions that resemble those that Gauthier defends as rational. This is significant for two reasons. First, it results from a combination of the descriptive and normative questions in a way not found in current literature. It is not all that rare to see some appeal made to evolutionary processes in the philosophical literature, but little attention has been paid thus far to the import of the cultural evolutionary story to moral theory in general and normative contract theory in particular.

Second, we will see that there is a unique structure to the outcome of the cultural analysis of human sociality that blends particularly well with Gauthier’s defence of constrained maximization. The prevalence of cooperation, I will argue, depends on the presence of prosocial dispositions that call for a constraint on the pursuit of self-interest. Rationality calls for the formation of the disposition to constrained maximization, which likewise requires similar constraints on the pursuit of self-interest. These results set the stage for a further examination into the connection between reason, evolution and morality. I will end by pointing to some possible directions in which we can flesh out these equivalences more fully.

3.2 What Is Cooperation?

‘Cooperation’ has distinct meanings in biological and moral contexts and varying meaning within each context. As such, it will first be important to clarify what we mean by the term. In ordinary language cooperation refers to any coordinated,

mutually beneficial behaviour. But it is commonly used to mean something different in evolutionary biology and moral theory.

In evolutionary biology, the terms ‘cooperation’ and ‘altruism’ are generally used interchangeably, and understood to refer to behaviour that is costly to the individual and beneficial to others. Peter Richerson and Robert Boyd say that they ‘use the word *cooperation* to mean costly behavior performed by one individual that increases the payoff of others. This usage is typical in game theory, and common, but by no means universal, in evolutionary biology’ (Boyd and Richerson 2006, p. 454). This equates cooperation and altruism, and Elliott Sober and David Sloan Wilson explicitly endorse that equivalence. They say, ‘prevalent among game theorists is their use of the word *cooperation* rather than *altruism*... the word *cooperation* is used by evolutionary game theorists, presumably because it is easier to think of cooperation as a form of self-interest. The behavior is the same but it is labeled differently’ (Sober and Wilson 1998, p. 84). In the moral context, David Gauthier (1979) uses ‘cooperation’ to describe the behaviour required by a particular subset of morality, namely that of distributive justice. He distinguishes two parts of morality: ‘distributive justice’ and ‘acquisitive justice’. The first ‘constrains the modes of cooperation’; the second ‘constrains the baseline from which cooperation proceeds’. Our concern in this paper will be with the former part of morality having to do with the emergence and maintenance of cooperation and the question why rational individuals should cooperate.

Thus, in very general terms, ‘cooperation’ can be understood as a type of behaviour that involves some kind of constraint on individual interest, where ‘individual interest’ can be understood in terms of biological fitness or rational self-interest. I will use the terms ‘cooperation,’ ‘altruism,’ and ‘morality’ (i.e., the subset of morality Gauthier identifies) interchangeably. In doing this I blur finer distinctions between them. Nonetheless, this allows me to conveniently talk about the central problems of evolution and rationality outlined above, viz., ‘How is it possible for organisms to act in such a way as to lower their fitness?’ and ‘How is it possible for rational beings to act contrary to their self-interest?’

3.3 The Descriptive Problem

The first aim of this paper is to address the two problems of cooperation outlined above. This involves answering two questions. The first is why cooperative behaviour is so prevalent in nature. This is largely a descriptive or explanatory question. The second is whether rationality dictates that individuals ought to cooperate. This is primarily a justificatory question. With respect to the descriptive question, explanations of cooperation can be one of two types. The first is to explain the behaviour in terms of its proximate mechanisms. Explanations of this sort will usually appeal to the underlying psychological mechanisms responsible for that behaviour. The second explanation appeals to ultimate causes. Explanations of this sort will generally make reference to the ultimate evolutionary mechanisms that produce such behaviour.

We might leave a tip in a foreign restaurant because to do otherwise would leave one feeling guilty. Or we might help a stranger in need because their plight elicits in us pangs of empathy. These psychological affectations are the proximate mechanisms of said behaviour. The ultimate explanation, on the other hand, will appeal to the evolutionary benefits that having such dispositions has for the one who possesses them. We might say, for example, that being disposed to feeling guilty for transgressions of norms pertaining to, say, tipping in foreign restaurants, leaves one more biologically fit than another who is not so disposed.

Thus, explaining the emergence of cooperation requires addressing what psychological mechanisms underlie that behaviour, and what evolutionary processes produce those mechanisms. But providing these explanations is not straightforward. The central difficulty is to articulate the mechanisms required to support cooperative behaviour in a way that is amenable to what is known about how evolutionary processes work. Given both the general structure of cooperative behaviour – viz., that it imposes a cost on its actor and benefits the recipient – and of evolutionary processes that presumably require any costs associated with a particular behaviour to be recouped if that behaviour is to persist, there is a tension between the explanandum and explanans. Since cooperation involves an individual's sacrifice of his or her reproductive fitness in order to enhance that of another individual, and since natural selection works against fitness-decreasing characteristics, there is a difficulty in reconciling the two.

Two standard mechanisms invoked to explain cases of cooperative behaviour are kin selection and reciprocal altruism. Parents act in ways that appear to directly reduce their own fitness and promote the fitness of their offspring. Kin selection provides an explanation why. Since offspring contain on average one half of the genes carried by either parent, caring for offspring, which is itself disadvantageous to the individual, is a behaviour that promotes the survival of one's genes. We can invoke kin selection to explain other behaviours that appear detrimental to the one performing them. Some birds display warning calls, which renders the individual making the warning call more likely to be picked out by a predator than the birds who are warned. Again, we have here an instance of an apparently fitness-decreasing behaviour, leaving one to ask how that could have evolved. The answer is this. Birds who display warning calls will live in groups of closely related kin. While the individual who warns others may be more likely to fall prey to a predator, his call promises to save many of his kin who share his genes. Thus, this seemingly fitness-decreasing behaviour is rendered compatible with natural selection, since this behaviour is actually one that promotes the survival of one's genes.

Reciprocal altruism permits the same kind of explanation in cases where individuals are not related. Nit-picking in birds, grooming among chimpanzees, and sharing of blood among vampire bats are all examples of apparently altruistic behaviours. On closer inspection, we see that the apparent fitness costs imposed by these acts are recouped through reciprocity and can thus be made sense of within an evolutionary framework. But while it is fairly widely accepted that altruistic behaviour in non-human animals can be explained through the evolutionary mechanisms of kin selection and reciprocal altruism, it is not obvious that these mechanisms are

sufficient to explain the large-scale cooperation among non-related individuals common in human societies. Certainly in some cases kin selection and reciprocity come into play. But when individuals are unrelated, we can rule out kin selection. And as group size increases, it becomes more difficult to keep track of one's past cooperative partners, and thus reciprocal altruism no longer seems a plausible candidate to explain cooperation on large scales.

There are a number of more sophisticated models of reciprocity that ultimately hinge on 'enlightened self-interest' to explain the emergence of cooperation, but these too all fall short. Indirect reciprocity is one such mechanism, which relies on reputation to determine one's partners. Other proposed mechanisms, like costly signaling, involve the display of fitness-decreasing traits that signals prestige or strength, and consequently enhance individual reproductive advantage.¹ Thus, the cost of the signal is recouped by it granting greater access to mates. But mechanisms like these that appeal to individual advantage are sometimes difficult to square with observed behaviour. We will often do things such as leave a tip in a foreign restaurant, warn drivers that they left their car lights on, return found items, draw attention to being undercharged, and so on, where the preservation of reputation (or any kind of 'individual advantage') is not a plausible explanation. And as Richerson et al. ask: 'If a mechanism like indirect reciprocity works, why have not many social species used it to extend their range of cooperation?' (Richerson et al. 2003, p. 379). Individuals seem to perform actions that appear to be genuinely fitness-decreasing and we must now ask whether this behaviour can be rendered compatible with the standard tenets of evolutionary processes.

Sober and Wilson (1998) think that they can be, and appeal to group selection to explain how. According to them, individual-level explanations of the emergence of altruism require that the cost of altruism be recouped elsewhere and, thus, fail to provide an explanation of *genuine* altruism. That is, parental care, or helping one's neighbour, are really instances of selfish behaviour: these are instances of behaviour that benefit the individuals themselves or their genes. Sober and Wilson contend that altruism – *genuine* altruism, i.e., behaviour done for others that lowers the individual's fitness without genetic or other recouping – can be explained in terms of group selection. According to the group selection hypothesis, some behaviours or traits evolved, not because they were advantageous to particular individuals, but because members of groups containing those traits did better than members of groups that did not. While individual selection will favour the evolution of selfishness within groups, and thus altruists will be less fit relative to non-altruists within a single group, matters are different at the level of groups. Altruists will do, on average, worse than selfish individuals within the same group. But, so the argument goes, members of groups of altruists will do better than will members of selfish groups. And if we grant that selection can occur at the level of groups, then

¹An example of this is the peacock's tail: the ornate tail is certainly a hindrance but it does serve as a signal that the peacock is strong and healthy enough to survive and preserve its nice tail and, thus, is likely a good selection for a mate.

we can explain the existence of altruism in terms of it: altruism evolved because members of altruistic groups did better than members of non-altruistic groups. Group selection thus promises an explanation of the emergence of behaviours or traits that genuinely reduce the fitness of an individual within a particular group, so long as those behaviours or traits increase the fitness of the members of the group that contains them relative to members of groups that do not.

But how plausible is this? Let us say that for every recipient of an altruistic act, an individual gains 2 Darwinian fitness points. Every altruist loses 1 fitness point for every act of altruism she performs. When two altruists meet, each gains 2 but loses 1. When an altruist meets a selfish individual, she loses 1 and confers a benefit of 2 onto the selfish individual. When two selfish individuals meet, neither gains and each receives a payoff of 0. Of two populations, one containing all altruists, the other containing all selfish individuals, the population of altruists will be fitter than the selfish group, since altruists each receive 1 point when they encounter one another, and selfish individuals receive nothing. If we translate these fitness points to number of offspring, we can see that the proportion of altruists in the global population (viz., the population resulting from the combination of the two groups) will increase.

However, in a population where both altruists and selfish individuals are present, altruists will be at a substantial selective disadvantage. In groups containing both altruists and selfish individuals, even if altruists make members of the group that contains them more fit than members of the group that does not, within the group, altruists will be less fit than non-altruists. Selfish individuals will have higher levels of relative fitness than will altruists of the same group, and will thus have more offspring than altruists. Selfish individuals will prosper, and the number of altruists within the population will decline.

Let us suppose that a population consists of two individuals, Anne and Sam. Anne is an altruist; Sam is a selfish type. As an altruist, Anne will behave in ways that reduce her own fitness and in ways that promote the fitness of those around her. Consequently, Anne will have only one offspring (let us suppose she does so asexually); while the selfish recipients of her altruistic acts will have two. Sam, as a selfish type, will not engage in any fitness-detrimental behaviour but will benefit from the altruistic actions of others. Sam will have two offspring. In the next generation, Anne's offspring, as an altruist like her mother, will have only one offspring. Sam's two offspring, on the other hand, will each have two offspring. Thus, after one generation the population will consist of three altruists (Anne, her offspring, and her offspring's offspring), and seven selfish types (Sam, Sam's two offspring, and two offspring each of Sam's two offspring). In the next generation, the number of altruists relative to selfish types will decline even further. And so on. Thus it appears that so long as selfish types are present in a population, altruists should tend towards extinction. Populations ought to contain very few altruists relative to selfish individuals. But this is not the case, and invites the question why not.

Sober and Wilson appeal to Simpson's Paradox to explain how the evolution of altruism in mixed populations like the above is possible. Simpson's Paradox 'refers to the phenomenon whereby an event *C* increases the probability of *E*

in a given population p and, at the same time, decreases the probability of E in every subpopulation of p (Pearl 2000, p. 1). Sober and Wilson illustrate Simpson's Paradox with an example of a discrimination inquiry at the University of California Berkeley.² Based on the smaller overall percentage of women who were admitted than that of men, it was suggested that the University's admission policies were discriminatory. Upon further investigation, however, it was discovered that every department admitted an equal proportion of women to men. And yet fewer women, overall, were admitted than men.

This seemingly paradoxical result can be explained as follows. It was discovered that a greater number of women tended to apply to departments that had lower acceptance rates than those to which men tended to apply. Let us say that department A accepts only 25 % of its applicants, while department B accepts 75 %. Suppose that department A receives the following distribution of applicants: 80 women and 20 men. Department A accepts 25 % of the women applicants and 25 % of the men applicants: 20 women and 5 men. Suppose now that 20 women and 80 men apply to department B. Since department B accepts 75 % of their applicants, the result is that 15 women and 60 men are accepted. A combined total of 100 men and 100 women applied to the two departments. Only 35 women in total were admitted, while 65 men were admitted. And yet, no department was discriminatory in their acceptance rates, since each department accepted an equal proportion of men to women.

Sober and Wilson employ Simpson's Paradox to reveal that it doesn't necessarily follow that, based on the fact that altruists will be less fit in a subgroup than selfish individuals, they will necessarily be less fit overall.³ While selfish individuals may do better within groups than altruists, if two groups are combined, the reverse effect can be obtained. Thus, 'what is true, by definition, is that altruists are less fit than selfish individuals *in the same group* . . . however, nothing follows from this as to whether altruists have lower fitness when one averages *across all groups*' (Sober and Wilson 2000, pp. 190–191).

We thus get a sketch of the kind of contribution group selection can make to explaining the emergence of traits that appear to perform to the detriment of their possessors and to the advantage of the group in which they occur. Individual selection may favour the evolution of selfishness within groups (and thus altruists will be less fit relative to non-altruists within a single group), but members of altruistic groups will do better than members of selfish groups.

²This example has its origin in Cartwright (1979), pp. 419–437.

³There is some controversy over how to interpret Simpson's Paradox and also whether Sober and Wilson employ it appropriately for their purposes. A fuller discussion of these issues lies beyond the scope of my aim here. According to Nancy Cartwright (1979) and Judea Pearl (2000), it is the fact that there is a correlation between being a woman and applying to departments with lower acceptance rates that is responsible for the statistical anomaly in the Berkeley case. Pearl says: 'In the case of Simpson's paradox, we have a clash (i) between the assumption that causal relationships are governed by laws of probability calculus and (ii) the set of implicit assumptions that drive our causal intuitions' (Pearl 2000, p. 7). It's not clear that Sober and Wilson recognize this. On this, see further: Peter Gildenhuys (2003), p. 36.

The concept of group selection is not new and was invoked by Darwin himself to explain human moral behaviour:

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and advancement in the standard of morality will certainly give an immense advantage to one tribe over another. (Darwin 1871, p. 166)

But while the promise to explain the emergence of genuinely altruistic behaviours is attractive, the plausibility of the theory is highly contentious. Perhaps the most pressing problem facing group selection is the unlikelihood of the conditions required to make it work. Sober and Wilson maintain that the evolution of altruism is an outcome of a conflict between two competing processes: individual selection within groups and group selection between groups. For altruism to evolve by group selection, altruists must be distributed in populations in such a way that groups containing altruists will increase at a more rapid rate than groups that contain selfish individuals and, thus, the global population of altruists will grow in spite of their being disadvantaged at the individual level. Maynard Smith (1964) produced the Haystack model, which modelled these conditions. They are as follows. (1) Groups must be isolated and sufficiently varied; otherwise the effects of group selection will be negated. (2) Groups must divide and intermix at just the right time to permit differential reproduction of altruism and selfishness and also to prevent individual selection from driving altruism to extinction. (3) Groups must then re-isolate themselves, and the process must then be repeated. Maynard Smith was sceptical that these conditions would actually obtain in nature, and thus that group selection played much of a role in the evolution of altruistic tendencies. More specifically, what is missing from the biological account is the plausibility that genetic differences will be distributed in such a way that there is sufficient variation between groups and sufficient similarity within groups.

The evolution of altruism by group selection thus requires that altruists be aggregated within and between groups in a way that is unlikely to occur at the biological level in human groups. As Richerson and Boyd say, ‘The trouble with a straightforward group selection hypothesis is our mating system. We do not build up the concentrations of intrademic relatedness like social insects, and few demic boundaries are without considerable intermarriage’ (Richerson and Boyd 1998, p. 81). And further, ‘Even very small amounts of migration are sufficient to reduce the genetic variation between groups to such a low level that group selection is not important’ (Richerson and Boyd 2006, p. 463). If so, to the extent to which variation within and similarity between biological groups is needed for altruism to evolve, biological group selection is unlikely to have played a significant role in the evolution of altruism in humans.

But while group selection at the biological level is unlikely a strong evolutionary force due to the implausibility that genetic differences will be distributed in such a way that there is sufficient variation between groups and sufficient similarity within groups, this does not hold in the case of cultural variants. Richerson and Boyd contend that cultural variants in the form of social norms will help to generate

suitable variation between groups and homogeneity within groups to permit group selection at the cultural level. Some groups will operate with more successful norms than others and, thus, will out-compete groups operating with less successful norms, and as those groups flourish so will their norms.

According to Richerson and Boyd, ‘Culture is information capable of affecting individuals’ behaviour that they acquire from other members of their species through teaching, imitation, and other forms of social transmission’ (Richerson and Boyd 2005, p. 5). This can range over particular beliefs (e.g., belief in God), skills (e.g., tool use, technological innovations), behaviours (e.g., washing food prior to eating it, removing one’s hat when indoors), or social norms (e.g., pay taxes, obey water restrictions, take only your fair share of resources, etc.). Richerson and Boyd argue that cultural adaptation has resulted in significant behavioural differences between groups. Culture permits individuals to adapt quickly to environmental changes. Consequently, individuals are very locally adapted to a wide range of environments, which has resulted in significant variations in behaviours between human groups. And the conformist tendencies of humans, together with intolerance of differences, tend to keep groups uniform. These differences between groups, together with competition between groups, set the stage for group selection.

Group selection results in the selection of group-beneficial traits. Groups that operate more advantageous norms will out-compete those groups that operate less advantageous norms. Thus when groups with different cultures with differential fitnesses come into conflict, and one wins out over the other, the culture of the winning group grows, and the losing group either is extinguished or absorbed. On their own, these norms might permit societies operating with them to function satisfactorily. But when they are operating within the larger structure in the world, they are being outcompeted by other social structures. And Richerson and Boyd, following Darwin, think that cooperative groups will tend to out-compete non-cooperative groups, and thus a cooperative culture will take root and grow.

Thus, cultural evolution generates differences between groups. These differences make for an environment conducive to group selection. More cooperative groups do better than non-cooperative groups. And since, according to Richerson and Boyd, genes and culture co-evolve (Richerson and Boyd 2005, pp. 191–236), and since a cooperative culture will make groups more successful than groups that do not employ cooperative norms, individuals who have dispositions towards such cooperative norms will do better than individuals without them. An environment is thus created that is conducive to prosocial dispositions. These include an ability to internalize and conform to norms and the capacity for feelings of guilt and shame, and are dispositions that increase the chance that norms are followed. The emergence of prosocial dispositions thus feeds back into the support and maintenance of cooperation.

We now have a promising solution to our first puzzle: group selection on cultural variants provides us with a plausible evolutionary account of the existence of cooperation in human beings. This, however, is not the only possible solution. Its rival is the so-called ‘Big Mistake Hypothesis.’ On this view, the social dispositions that we have are left over from earlier times when human groups were composed

largely of closely related kin. On this view, our psychological dispositions towards cooperative behaviour are once-adaptive responses to an environment in which we no longer find ourselves, and evolution hasn't had a chance to adapt to modern social group composition. The widespread cooperation we find is thus a 'Big Mistake'.

The Big Mistake Hypothesis is a coherent story and does appear to explain widespread cooperation that extends beyond immediate kin ties and reciprocal relationships. Confirming or denying the validity of this hypothesis requires in part a fuller investigation into the history of human sociality to uncover at what point early human social structures began to diverge from kin-based clan and at what point maladaptive traits (like prosocial dispositions, according to the Big Mistake Hypothesis) emerged (Richerson and Boyd 2005, pp. 188–189). Evaluating the success of this account lies beyond the scope of my aim here and is unsettled in the literature. But even if the Big Mistake Hypothesis were true, that fact alone would not preclude that culture has played a significant role in our evolutionary history and shaped at least part of human social behaviour.

3.4 The Normative Problem

I now move on to the second problem of cooperation, namely reconciling cooperation and rationality. David Gauthier has claimed that, 'the reconciliation of morality and rationality is the central problem of modern moral philosophy' (Gauthier 1990, p. 150). Given that rationality requires the pursuit of one's self-interest, and morality (or, in our case, cooperation) constrains the pursuit of individual interests, it seems that moral behaviour is irrational.

The Prisoner's Dilemma provides a formalization of the problem of cooperation. It involves two accomplices who are caught for committing a crime, interrogated separately, and offered a deal. If one player incriminates the other, or 'defects', while the second remains silent, or 'cooperates', he will be given a sentence of 1 year, while the other player will get four. If both remain silent, both will be sentenced to 2 years, but if both defect, both will receive 3 years. The following matrix represents this game.

		Prisoner 2	
		Cooperate	Defect
Prisoner 1	Cooperate	2 years, 2 years	4 years, 1 year
	Defect	1 year, 4 years	3 years, 3 years

The first number of each pair represents Prisoner 1's possible outcomes; the second number Prisoner 2's. In this particular case, no matter what the other player does, defecting is the utility-maximizing response. However, given that each player

is rational, both will employ this equilibrium strategy, which will lead to a situation that is less preferred to the one where both cooperate. Rationality thus sometimes leads players to a suboptimal outcome.

Those who think cooperation can be reconciled with rationality will point to the fact that mutual cooperation will yield a higher utility than will mutual defection. But proponents of a strictly maximizing conception of rationality will contend that in the Prisoner's Dilemma to cooperate is a dominated strategy (that is, no matter what one's opponent does, defection is always the best reply in terms of utility maximization) and is thus positively irrational.

Hobbes's Foole takes this line, and asks why one cannot violate the rules of morality in cases where doing so is advantageous.

The Foole hath sayd in his heart, there is no such thing as Justice; and sometimes also with his tongue; seriously alleaging, that every mans conservation, and contentment, being committed to his own care, there could be no reason, why every man might not do what he thought conduced thereunto: and therefore also to make, or not make; keep, or not keep Covenants, was not against Reason, when it conduced to ones benefit. (Hobbes 1651, ch. 15, par. 4, 74)

The Foole's objection points to a structural problem in the Prisoner's Dilemma: it will always be to one's benefit to violate one's agreement. And insofar as one is rational to the extent that one pursues one's benefit, the rational course of action will always be to defect.

In reply to the Foole, Hobbes argued that violations of morality are liable to be detected and punished, and that the consequences of being caught – that is, being excluded from civil society – are so grave that defection is never a good gamble. This reply in effect changes the payoff structure of the Prisoner's Dilemma such that unilateral defection carries with it grave consequences rather than rich rewards. But in order for this reply to be successful, the highly unlikely state of affairs must obtain where, for every single potential defection, the risks and possible costs associated with defection outweigh any possible gains. Hobbes's ultimate solution to the problem of non-compliance is a political one: to have a sovereign with sufficient power of surveillance and authority to punish so as to make non-compliance counterproductive. This, however, can be costly and inefficient, and it would be desirable if compliance could be achieved by more efficient non-coercive means.

Gauthier presents us with such a means. He locates the rationality of compliance with agreements in the adoption of the disposition he refers to as 'constrained maximization'. Constrained maximizers conditionally dispose themselves to cooperation. This disposition to cooperate distinguishes the constrained maximizer from what Gauthier refers to as a 'straightforward maximizer', who 'seeks to maximize his utility given the strategies of those with whom he interacts' (Gauthier 1986, p. 167). Straightforward maximizers (like the Foole) are rational utility-maximizers; they will defect when it is advantageous for them to do so. The constrained maximizer, by contrast, will cooperate when he expects others to cooperate and defect only when he anticipates that others will do the same.

The underlying presupposition of constrained maximization is the rationality of constraining utility-maximization in order to gain a mutually optimal strategy. The constrained maximizer disposes himself, essentially, to forgo token opportunities to make big gains through defection in order to obtain the benefits of mutual cooperation. Gauthier must then show that doing so is rational. This requires showing that the gains through mutual cooperation outweigh gains through defection, a conclusion that is not obvious given the structure of interactions between the two types of maximizers.

In a society composed of a mix of straightforward and constrained maximizers, individuals will defect in all but two cases. First, two constrained maximizers interact and are able to identify each other as constrained maximizers. In such a case, both will cooperate. Second, a constrained maximizer mistakes a straightforward maximizer for a constrained maximizer. In such a case, the constrained maximizer will adhere to the agreement, and the straightforward maximizer will defect.

In Prisoner's Dilemmas the most utility-maximizing strategy is unilateral defection. If the straightforward maximizer is able to fool the constrained maximizer into thinking that he is a constrained maximizer, then he will be able to gain at the expense of the constrained maximizer. This will yield the best outcome for the straightforward maximizer and the worst for the constrained maximizer. The second best outcome for both is mutual adherence, and the third best for both is mutual defection.

When neither party cooperates, both the straightforward and the constrained maximizer will receive the third best payoff. In an encounter between a straightforward maximizer and a constrained maximizer, if the constrained maximizer adheres to the agreement while the straightforward maximizer defects, the straightforward maximizer will receive a payoff greater than he would had he adhered. The straightforward maximizer is thus able to reap advantages unavailable to the constrained maximizer through unilateral defection. Although the constrained maximizer has available to him opportunities for gain through mutual adherence, mutual adherence yields a utility less than does unilateral defection.

In order to avoid the impending conclusion that rational individuals ought always to defect, Gauthier must rely on the assumption that straightforward maximizers cannot pass as constrained maximizers. For, given that straightforward maximizers will do better than constrained maximizers as long as they are given the same opportunities as constrained maximizers, and whether one is able to partake in agreements with others will depend on whether it appears that one can be trusted, so long as one is able to maintain the illusion of being a constrained maximizer, one will be able to partake in agreements with others while reaping the benefits of defection. Straightforward maximizers thus will do better than constrained maximizers, which will prevent one from claiming that constrained maximization is rational.

Gauthier recognizes how crucial the detection of dispositions and intentions is. If people were (to use Gauthier's terminology) *transparent* (i.e., their characters were always accurately detectable by others), then constrained maximization would

be the rational strategy. For in that case constrained maximizers would be able to identify straightforward maximizers, and consequently exclude them from agreements. If people were, on the other hand, what Gauthier describes as *opaque* (i.e., their characters remained hidden to others), then straightforward maximizers would do better. For straightforward maximizers would then be able to continue to make agreements with others and gain through successful exploitation. Gauthier argues that neither of these is realistic, and claims that people are *translucent*, according to which ‘persons are neither transparent nor opaque, so that their disposition to co-operate or not may be ascertained by others, not with certainty, but as more than mere guesswork’ (Gauthier 1986, p. 174). In other words, if our characters are translucent, it is assumed that one has a better chance of correctly identifying another’s character than one would by randomly guessing.

If so, and if one who is suspected to be a straightforward maximizer will be excluded from cooperative agreements, then the straightforward maximizer will not reap all projected benefits of both cooperation and defection. The straightforward maximizer will forgo many benefits of cooperation and reap only those gains that he can through successful exploitation. As his untrustworthy character becomes more widely known, opportunities for exploitation will diminish. And since constrained maximization affords one the possibility of gaining through mutually-beneficial cooperative interactions – opportunities unavailable to the straightforward maximizer – it becomes plausible to suggest that constrained maximization will yield a higher utility to those who so dispose themselves than will straightforward maximization. Gauthier thus concludes that rational persons will become constrained maximizers (Gauthier 1986, p. 128). And if Gauthier’s defence of constrained maximization is successful, then he will have provided an account that reconciles cooperative behaviour with rationality.

The success of Gauthier’s reconciliation project depends primarily on the answers to two questions. The first is whether constrained maximization is rational. The second is whether acting on that disposition is rational.

In order to establish the rationality of constrained maximization, Gauthier must argue that those who develop that disposition will do better than those who do not. This rests primarily on the plausibility of translucency. This seems to be largely an empirical matter. There is evidence on both sides. Gauthier could point to physiological reactions that accompany deceit – rapid heartbeat, flushing, aversion of eye contact, and so on, to support the view of the detectability of dispositions. He could also say that we can determine the dispositions of others in more impersonal ways by examining the history of their behaviour. McClennen writes:

As it turns out, the iterated game framework provides a setting in which the epistemological problem of assurance can be resolved. If interaction is sufficiently on-going, then for many kinds of encounters, a given individual can have the requisite assurance regarding the disposition of other participants. The history of past encounters between participants typically will provide the needed information. It is plausible to suppose, moreover, that for many such contexts, the requisite information will be securable from anecdotal sources – that is, it will be unnecessary to resort to formal mechanisms for the compiling and transmission of this information. At the ‘street level’, each typically will be able to

consult personal experience and informally shared information with friends and family members to determine whether or not the level of voluntary cooperation in more impersonal, 'public' settings has been great enough to warrant voluntary compliance on one's own part. (McClennen 2001, p. 203)

Critics will say otherwise, and may point to instances where individuals successfully lie or cheat their fellow men. The issue is yet to be settled, but as it stands, there is no real knockdown argument against translucency, and I thus conclude that Gauthier's argument for constrained maximization remains undefeated on the ground of the implausibility of translucency. Furthermore, there is reason to believe that the evolutionary story of the origins of human cooperation can help to bolster the view that humans have evolved as translucent creatures. If the story I have endorsed about group selection on cooperative cultural variants and individual selection of prosocial dispositions is defensible, then it would be plausible to suggest that among those prosocial dispositions would have also evolved an ability to detect the character of others and signal one's own trustworthy character to others.

It has been suggested that even if it can be established that those who develop a disposition to constrained maximization do better than those who do not, that does not entail that it is also rational to act on that disposition (Kavka 1987; Parfit 2001). In other words, the success of Gauthier's project rests not only on whether it is rational to acquire the disposition to constrained maximization, it must also be established that it is rational to act on that disposition. Critics of Gauthier might grant that it is rational to form a disposition to constrained maximization but deny that it is rational to carry through with that disposition. Consider Gauthier's example of harvesting one's crops, borrowed from Hume (Gauthier 1994, p. 692; Hume 1988, Book iii, Part ii, Section iv, pp. 520–521). Persons A and B have crops to harvest. They can do so alone or they can agree to help one another. Helping one another will involve one person helping the other at T1, and the other returning the help at T2. In order for person A to receive assistance from person B in harvesting her own crops, she will have to assure B that, after B helps her at T1, she will provide B with assistance at T2. Suppose that B helps A at T1. Is it now, at T2, rational for A to help B?

Critics of Gauthier will contend that it is not. Their argument might go like this. A's act of assuring B constitutes A forming a disposition to help B. Assuming translucency, A's forming of the disposition causes B to help A. A then gains from the disposition – a gain she would not have obtained had she not formed the disposition. It is thus rational for A to form the disposition to help B. But now comes T2. Is it now rational for A to actually do what she disposed herself to do? Helping B at T2 imposes a cost on A. At T2 A has already received B's help and so (let us suppose) has nothing to further gain from helping B at T2. Thus, A has no reason to help B at T2. How can we say that helping B is now rational?

To relate this example to constrained maximization, while it might be rational to dispose oneself to constrained maximization at T1, insofar as doing so secures a cooperative gain at T2, it is not clear that at T2 carrying through with one's part of the agreement is rational. At T2, having already secured the cooperative outcome, the individual disposed to constrained maximization has nothing further to gain

from that disposition. At T2, it is utility-maximizing for an individual to act as a straightforward maximizer. Thus it would seem rational to, at T2, abandon actions recommended by constrained maximization.

Gauthier resists this conclusion and argues that the rationality of forming the disposition to constrained maximization does indeed entail the rationality of actions recommended by that disposition. According to him, there is a crucial relationship between an agent's assurance that she will carry through with her agreement, her intention to do so, and her success in securing the cooperative outcome. Specifically, Gauthier contends that in order for A to receive the cooperative benefit from B at T2, A must provide B at T1 with a sincere assurance that she will carry through with the agreement. In order to provide a sincere assurance at T1, A must have the intention that at T2 she will carry through with the agreement. In order to have the intention to cooperate at T2, A must believe that it is rational for her to do so. According to Gauthier, A cannot, without inconsistency, provide B with sincere assurance that she will cooperate at T2 if she knows that at T2 it will no longer be to her benefit (and will thus be irrational) to cooperate.

But if A evaluates the rationality of her actions at the level of action, she must concede that it is not rational to cooperate at T2, since at T2 cooperating is not the utility-maximizing option. If A evaluates her action this way, then she cannot make a sincere assurance at T1 to cooperate at T2, since she knows that at T2 cooperating is not the utility-maximizing (and thus rational) option. Without the sincere assurance that A will cooperate, B will not either. Both A and B will end up with the non-cooperative outcome, and each will harvest her crops alone.

To avoid this outcome, Gauthier recommends the adoption of a different perspective from which to evaluate the rationality of action. His concern shifts to the rationality of dispositions rather than actions. According to Gauthier, 'intentional structures create problems for the orthodox account of deliberation, which insists that rational actions are those that directly promote the agent's aim, taking as illustrative the aim that one's life go as well as possible.' He continues, 'If my aim is that my life go as well as possible, then I should not take all of my reasons for acting directly from that aim, considering only which action will have best consequences for my life. For if I always deliberate in this way, then my life will not go best for me' (Gauthier 1994, p. 692).

Gauthier's revised account of rational deliberation permits A to cooperate at T2 so long as making the assurance at T1 to cooperate at T2 yields a greater utility than doing otherwise. And, 'since the direct link between rational deliberation and particular outcomes has been severed, an action may be rational even though at the time of performance it is not, and is not believed to be, part of a life that goes best for the agent' (Gauthier 1994, p. 171). Thus, it is rational for A to cooperate at T2.

Thus, on the above account of rational deliberation, good reasons are those that lead to optimality. This reconceptualization of reason permits Gauthier to say that an action recommended by constrained maximization is rational, even if there is an alternative act that will yield a higher utility so long as the disposition to constrained maximization yields greater utility than does the disposition recommending the alternative action (in our case, the disposition to straightforward maximization).

Actions that, if taken as individual tokens are non-utility-maximizing are rational, so long as they are recommended by a disposition that it is rational to have. This account resolves the problem of compliance and ensures a superior outcome to one obtainable from a maximizing conception of rationality.

One might ask whether Gauthier is entitled to make this kind of reconceptualization of rationality. McClennen (2001, pp. 189–208) thinks so. On his view, while the cooperative outcome is a dominated and non-equilibrium point in the Prisoner's Dilemma, that fact alone should not exclude that outcome as a resolution to the problem. According to him, convergence can in fact occur at loci other than equilibrium points, and he thinks that Pareto-Optimality is one such locus.⁴ He considers games of pure coordination, i.e. games where there is no conflict of interest between players and the goal is merely to coordinate their strategies. Of these games, he says:

The appropriate concern for rational players... is to coordinate strategies so that the outcome will satisfy the Pareto condition. It is true, of course, that outcomes satisfying the Pareto condition satisfy the equilibrium condition. But from the perspective of the strategic problem that players face in a game of pure coordination, this additional property is purely accidental. That is, the equilibrium concept adds nothing that illuminates the nature of the deliberation that persons face in such game. In this context, it does no work. (McClennen 2001, p. 196)

On McClennen's view, since in coordination games it is not clear that the equilibrium concept itself plays much of a role in the determination of strategic action, and since instead much of the work is (and, according to him, ought to be from the point of view of rationality) done by Pareto-considerations, there is reason to suppose that the same might go for cooperation problems. This opens the door to an account like Gauthier's where reason is reconceptualized and actions are rational when recommended by rational dispositions, even if those actions are not individually utility-maximizing (and are, in fact, out of equilibrium in one-shot Prisoner's Dilemmas). We thus arrive at a reconciliation between cooperative dispositions, the actions they recommend, and rationality.

3.5 Connecting the Descriptive and the Normative

Thus far we have taken up two questions. The first is the descriptive question of how to explain the emergence of cooperation. The second is the normative question of why one should cooperate. These two are structurally similar. The constraints imposed by natural selection in the descriptive case and by rationality

⁴Gauthier defines this as follows: 'An (expected) outcome is optimal (or, more fully, Pareto-optimal) if and only if there is no possible outcome affording some person a greater utility and no person a lesser utility. Alternatively, an outcome is optimal if and only if any other outcome affording some person a greater utility also affords some other person a lesser utility' (Gauthier 1986, p. 76).

in the normative case make explaining or justifying cooperation difficult. How is it possible that cooperation evolved, given the workings of natural selection? How can we justify cooperation, given the self-serving conception of rationality? Regarding the former, I have argued that cultural group selection provides us with a plausible explanation of the emergence of the widespread cooperative behaviours among human beings. In the normative context, I have argued that David Gauthier's argument for the rationality of adopting the disposition of constrained maximization is a defensible route to reconciling cooperation with rationality.

In this section I will examine the relationship between the descriptive and normative projects. I will argue that the descriptive and normative projects are not only dependent on one another, but converge on the same outcome. This convergence comes from two independent lines of enquiry. Cultural group selection permits us to explain the emergence of behaviours that are genuinely fitness-decreasing at the individual level but that are beneficial to groups of individuals that display these behaviours. Explaining the emergence of cooperation requires us to shift our perspective from the individual to the level of groups. Individually, cooperation is more costly than selfishness. Collectively, cooperation pays.

A similar shift in perspective is required to justify the rationality of cooperation. Gauthier's project is to show that cooperative behaviour is rational, in spite of it being disadvantageous at particular instances. This parallels the explanation of cooperation in nature, which, as I have argued, evolves in spite of it being disadvantageous at the individual fitness level. Non-cooperation is utility-maximizing at the level of outcomes: if rationality is evaluated as a best-reply to one's partner's actions, non-cooperation will always be rational. But by shifting evaluation of rationality from outcomes to strategies, the cooperative – and superior – outcome can be achieved. This perspective permits us to rationally justify the constraints that morality requires. Thus, just as it is the case that, contrary to what we might expect, evolution supports cooperation, so too is it the case that, contrary to appearances, rationality supports cooperation. There is an advantage to cooperative behaviour in a particular context. That is, when others cooperate and when cooperation permits out-competition of other groups in the evolutionary context, and when it permits a mutually preferred outcome to universal defection in the moral context, then cooperating is better than not cooperating.

If I am right that the correct explanation of the existence of cooperation in nature appeals to group selection on cultural variants, then we also are able to arrive at an evolutionary story of the emergence of certain prosocial dispositions, viz., ones that dispose us to comply with norms and agreements. Group selection helps to shape an environment that favours prosocial psychological mechanisms. As Richerson and Boyd put it, 'if generally cooperative behavior is favored in most social environments, selection may favor genetically transmitted social instincts that predispose people to cooperate and identify within larger social groupings' (Richerson and Boyd 2005, p. 215). These emotions, in turn, help to reinforce cooperative behaviour. As Bowles and Gintis suggest:

Some prosocial emotions, including shame, guilt, empathy, and sensitivity to social sanction, induce agents to undertake constructive social interactions; others, such as the

desire to punish norm violators, reduce free riding when the prosocial emotions fail to induce sufficiently cooperative behavior in some fraction of members of the social group. (Bowles and Gintis 2003, p. 433)

Prosocial dispositions are ones that, if evaluated at individual instances, are not fitness maximizing. And insofar as this is true, they are also structurally similar to those actions required by constrained maximization. And if Gauthier is right that the actions recommended by constrained maximization are rationally defensible, then we will have shown that the dispositions emerging from the evolutionary story that I endorse can also be rationally defended. Cultural evolution yields prosocial dispositions, which constrain our self-interested pursuits. In a similar manner, rationality requires that we form dispositions towards constrained maximization. Thus we have a structural coincidence between what evolution produces and reason dictates.

The structural similarity in the outcomes of both projects provides a mutual support for each. The prosocial emotions that emerge from the cultural story also help to fill in the contingent facts upon which the rationality of constrained maximization depends. The cultural story provides evidence for the existence of a population structure where constrained maximization is rational. And Gauthier's analysis of the rationality of actions recommended by rational dispositions suggests the possibility of a rational justification for the dispositions that evolution produces.

3.6 Implications of the Convergence

I will end by pointing to some further implications of the convergence that I have illustrated above. The prosocial dispositions that emerge from the cultural evolutionary story suggest that mere considerations of self-interest do not exhaust the list of reasons that agents will employ. Gene-cultural co-evolutionary theory gives an account of a genetic component of cooperative tendencies, and is able to tell a story of how it might be the case that certain genetically regulated dispositions (e.g., empathy, conscientiousness, etc.) might have evolved in the first place. Put roughly, the theory states that these dispositions emerged as adaptations to an environment where cooperative behaviours were advantageous. Thus, genetic dispositions to behave in 'appropriate ways' were selected in this new, culturally created 'cooperative' environment.

The presence of these prosocial dispositions points to the possibility of a reorientation of rationality and a corresponding naturalistic normative moral theory. Developing that lies beyond the scope of my aim here. However, the promise is this: an evolutionarily-informed reconceptualization of rationality would stretch the standard conception of self-interest to include those preferences endowed to us by nature, such as sharing, group membership, fairness, and so on. On this suggestion, included among those things that we want to maximize will be fairness, sociality, etc.

Such a revision and broadening of what counts as an interest in light of evolutionary theory promises to resolve some of the tensions of a Gauthier-type analysis with respect to including the vulnerable in the sphere of moral concern. The notion of cooperation for mutual advantage that lies at the heart of contractarian moral theory entails that those who are unable to contribute anything to the cooperative outcome – such as the mentally underdeveloped, infants, animals, and so on – are thereby excluded from the sphere of moral concern. This failure to extend moral consideration to non-contracting parties is a pressing problem for the contractarian. Gauthier handles this problem by restricting the scope of morality with which he is concerned. But by introducing a wider conception of interests, which can include things such as care for others, inclusiveness, and so on, we can accommodate those to whom we think concern ought to be granted without having to either narrow the scope of morality or abandon rationality. An evolutionary perspective supports a broadening of interests and the resulting conception of rationality, I contend, provides a more suitable basis for normative moral theory.

There are also some significant practical applications of an evolutionary account of cooperation, in particular with respect to institutional design. Institutional design has typically been modelled on the *homo economicus* view of human beings, and heavily governed by rules, monitoring, incentives and disincentives. However, the cultural evolutionary analysis of human cooperation supports a human motivation structure according to which we are not only driven by self-interest, but can also be altruistic. If so, the incentive systems that go with the *homo economicus* view can be replaced or supplemented with ones that stress other values such as fairness, autonomy, achievement, and teamwork. Such systems provide a more efficient and effective means of generating and sustaining cooperation, as evidenced by their success in other fields such as the automotive industry and software development. These less authoritarian frameworks promise to be generalizable to the development of policies and institutional design on larger scales, and to produce better outcomes for all.⁵

Acknowledgments Much gratitude is owed to my dissertation committee at the University of Toronto: Paul Thompson, Wayne Sumner, and Denis Walsh. I am very thankful for their comments on the dissertation, from which this paper is derived, and on earlier versions of the paper itself.

I am also grateful to my father and fellow philosopher, Alister Browne, who has patiently read and commented on everything I have ever written, and provided much needed support along the way.

References

- Benkler, Y. 2011. *The penguin and the Leviathan*. New York: Crown Business.
- Bowles, S., and H. Gintis. 2003. Origins of human cooperation. In *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, 429–443. Boston: Massachusetts Institute of Technology Press.

⁵For a development of this line, see Benkler (2011).

- Boyd, R., and P.J. Richerson. 2006. Culture and the evolution of the human social instincts. In *Roots of human sociality*, ed. N.J. Enfield and C. Levinson, 453–477. Oxford: Berg Publishers.
- Cartwright, N. 1979. Causal laws and effective strategies. *Nous* 14(4): 419–437.
- Darwin, C. 1871. *The descent of man and selection in relation to sex*. London: Murray.
- Gauthier, D. 1979. Bargaining our way into morality: A do-it-yourself primer. *Philosophic Exchange* 2(5): 14–27.
- Gauthier, D. 1986. *Morals by agreement*. Oxford: Oxford University Press.
- Gauthier, D. 1990. Justice and natural endowment: Toward a critique of Rawls's ideological framework. In *Moral dealings: Contract, ethics, and reason*, ed. D. Gauthier, 150–170. Ithaca: Cornell University Press.
- Gauthier, D. 1994. Assure and threaten. *Ethics* 104(4): 690–721.
- Gildenhuis, P. 2003. The evolution of altruism: The Sober/Wilson model. *Philosophy of Science* 70: 27–48.
- Hobbes, T. 1651. *Leviathan*. London: Murray.
- Hume, D. 1988. *A treatise of human nature*, ed. L.A. Selby-Bigge. Oxford: Clarendon.
- Kavka, G. 1987. Review: Morals by agreement. *Mind* 96(381): 117–121.
- Maynard Smith, J. 1964. Group selection and kin selection. *Nature* 201: 1145–1146.
- McClennen, E. 2001. The strategy of cooperation. In *Practical rationality and preference*, ed. A. Ripstein and C. Morris, 189–208. Cambridge: Cambridge University Press.
- Parfit, D. 2001. Bombs and coconuts, or rational irrationality. In *Practical rationality and preference*, ed. A. Ripstein and C. Morris, 81–97. Cambridge: Cambridge University Press.
- Pearl, J. 2000. *Simpson's paradox: An anatomy*. Cognitive Systems Laboratory, Computer Science Department, UCLA. Retrieved from: <http://escholarship.org/uc/item/8gh5613r>. Accessed 13 Oct 2011.
- Richerson, P.J., and R. Boyd. 1998. The evolution of human ultra-sociality. In *Ideology, warfare, and indoctrinability*, eds. I. Eibl-Eibesfeldt and F. Salter, 71–95. New York: Berghan Books.
- Richerson, P.J., and R. Boyd. 2005. *Not by genes alone*. Chicago: University of Chicago Press.
- Richerson, P.J., R. Boyd, and J. Henrich. 2003. Cultural evolution of human cooperation. In *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, 357–388. Boston: Massachusetts Institute of Technology Press.
- Sober, E., and D.S. Wilson. 2000. Summary of Unto others: The evolution and psychology of unselfish behavior. In *Evolutionary origins of morality*, ed. L.D. Katz, 185–206. Bowling Green: Imprint Academic.
- Sober, E., and D.S. Wilson. 1998. *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.