Egoism, Altruism, and Our Cooperative Social Order

Gerald Gaus

1 "Morality as Compromise"

As James Sterba recognizes in "Completing the Kantian Project: from Rationality to Equality,"[1] humans in society are often confronted by a conflict between self-regarding and other-regarding reasons. We are creatures devoted to our ends and concerns which give us self-regarding reasons, yet we are inherently social creatures and as such recognize other-regarding reasons. Morality, he claims is a way of commensurating these reasons. In particular, he argues for an "incomplete" conception of morality (57–8, 76) — "morality as compromise" — according to which (*i*) both altruistic and egoistic reasons are relevant to choice and (*ii*) while reason cannot necessarily provide the complete ranking of these two reasons, it can be rationally demonstrated that higher ranked altruistic reasons outweigh lower-ranked egoistic reasons. Many readers, no doubt, will be most interested in Sterba's striking claim that this "incomplete" conception of morality can be derived from very basic canons of good reasoning. Although I shall briefly address this ambitious thesis, my focus will be on where, if we accept it, it will take us. I wish to ask: if we accept Sterba's conception of morality as an incomplete compromise between egoism and altruism, how might we think about "completing" it — developing a more specific view of how a plausible human morality might "balance" egoism and altruism?

2 Two Orderings of Reasons

Basic to Sterba's argument is that both altruistic and self-interested reasons "are relevant to rational choice" (50–1). Our question when deciding what to do in some context is which of *these* two types of reasons should have "priority" (51). Sterba introduces the idea of two orderings or rankings of reasons: an altruistic ordering $\{a_1 \ldots a_n\}$ and a self-interested or egoistic ordering $\{e_1 \ldots e_n\}$ of reasons.[2] Let us suppose that each of these orderings satisfies the standard conditions of being (within each) complete and transitive. The question for a practically rational agent is how one is to combine these two partial orderings into an overall ordering of reasons to act.

The key move in his analysis is the contrast between three different ways of transforming these two partial (or sub-) orderings ($\{a_1 \ldots a_n\}$, $\{e_1 \ldots e_n\}$) into an overall ordering. Sterba compares three possibilities. First, $\{a_1 \ldots a_n\}$ may strictly dominate $\{e_1 \ldots e_n\}$ such that for any member a_i of $\{a_1 \ldots a_n\}$ and any member e_j of $\{e_1 \ldots e_n\}$, $a_i \succ e_j$ (i.e., a_i is preferred to e_j). Second, the egoistic ordering $\{e_1 \ldots e_n\}$ might strictly dominate $\{a_1 \ldots a_n\}$, such that $e_j \succ a_i$. The third possibility is that in the overall ordering neither sub-ordering strictly dominates the other, but, in the overall ordering, there are some altruistic reasons that are ranked higher than some self-interested, and some self-interested that are ranked higher than some altruistic. Sterba holds that the third option is "rationally required:"

> Once the conflict is described in this manner, the third solution can be seen to be the one that is rationally required. This is because the first and second solutions give exclusive priority to one class of relevant reasons over the other, *and only a question-begging justification can be given for such an exclusive priority*. Only by employing the third solution, and sometimes giving priority to self-interested reasons, and sometimes giving priority to altruistic reasons, can we avoid a question-begging resolution….Such a compromise would have to respect the rankings of self-interested and altruistic reasons imposed by the egoistic and altruistic perspectives, respectively. Accordingly, any nonarbitrary compromise among such reasons in seeking not to beg the question against either egoism or altruism would have to give priority to those reasons that rank highest

> in each category. Failure to give priority to the highest-ranking altruistic or self-interested reasons would, other things being equal, be contrary to reason (52; emphasis added, note omitted).

I cannot see any reason to think dominance solutions necessarily, as a matter of logic or basic principles of good reasoning, beg any questions. That we can order the set of considerations {$a_1 \ldots a_n$}, and that we can order the set {$e_1 \ldots e_n$} and that both are relevant to choice in context *C* does not tell us anything whatsoever about the combined ordering of *C*-relevant reasons (beyond that it should be consistent). To be able to show that the canons of reasoning themselves, apart from any substantive considerations, allow us to say something substantive about the features of the combined ordering would certainly be remarkable. Suppose one introduces a principle of set union of the following sort: *when combining two complete orderings it must be the case that one ordering cannot dominate the other in the complete ordering*. This would yield Streba's result, but surely *it* "begs the question" by eliminating all possibility of dominance solutions without considering the substantive case.

However, at least on one interpretation Sterba's core claim is so modest and reasonable that, even if we are suspicious of the strong claims about its rational necessity, we should still accept it. When he says that according to an acceptable "ordering, high-ranking self-interested reasons have priority over conflicting low-ranking altruistic reasons, other things being equal, and high-ranking altruistic reasons have priority over conflicting low-ranking self-interested reasons, other things being equal" (54), we might take him as simply saying that in the comprehensive ordering of reasons at least the very highest ranked reasons of one subordering ordering must ranked above the very lowest of the other. While that does not seem mandated by reason, if one admits that there are both egoistic and altruistic reasons, only an extraordinarily self-indulgent or self-sacrificial view would hold that in absolutely every case one type must trump the other. Instead of a simple ordering, think of the issue in terms of continuous values: surely in any plausible system of trade-off rates there must be some case in which the marginal importance of the self-interested (or altruistic) reason is so low that the value of acting on the other type is greater. I am not entirely confident that Sterba only has this minimalistic interpretation in mind; at some points there is the suggestion of *prima facie* equal importance.[3] *That* certainly would be a strong claim. In any event this raises the core question posed by morality as compromise: in the overall ordering how important are altruistic reasons, and how important are self-regarding ones?

3 Self-interest and Altruism in Human Society

3.1 Decision Functions

If principles of reason and rational argumentation do not tell us much about how we must compare self-interested reasons and reasons of altruism, how do human communities come to resolve these issues in their conceptions of morality? Let us continue with Sterba's supposition that we are confronted by two rankings of reasons, altruistic and self-interested. As he points out (52ff), the important cases for morality concern conflicts — when altruistic reasons point us one way, and self-interested ones another. What to do? Consider Figure 1, which depicts three decision rules as to how to resolve the conflict.

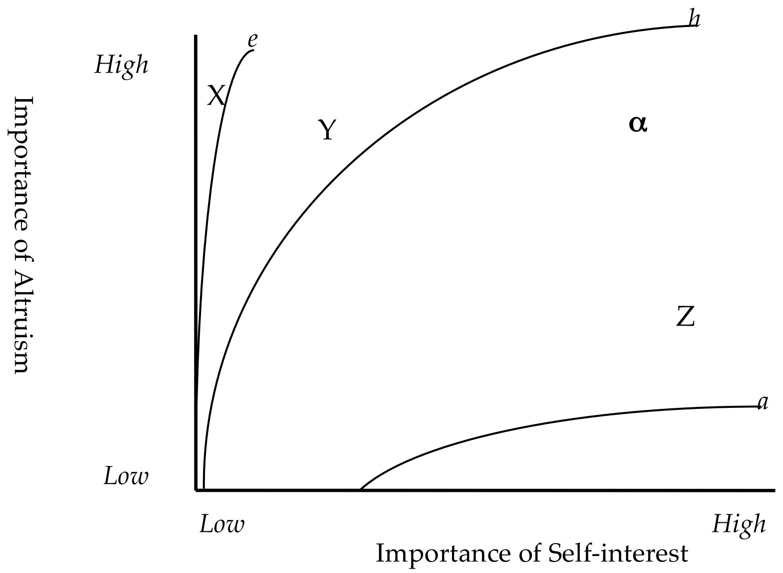

Figure 1

We can map any conflict situation between reasons of self-interest and altruism in this space; thus in case α the chooser is faced between acting on strong reasons of altruism and strong reasons of self-interest. Each curve (*e, h, a*) divides the decision space into essentially two parts: above and to the left of a curve a person will choose in a case of a conflict to act altruistically; below and to the right a person will act on self-interested reasons. (If a case falls exactly on a curve the person will be indifferent.) Curve *e* depicts a pretty thoroughgoing egoist; she will act on altruistic reasons only in cases in area X. Only when reasons of self-interest are very low, and altruistic concerns are very high, will she act on altruistic reasons. Curve *h* is a somewhat more altruistically-inclined trade-off rate, choosing altruistic reasons over self-interested ones in all cases in areas X *and* Y. Finally, curve *a* depicts the trade-off rate of a highly altruistic person, acting on altruistic reasons in areas X, Y and Z. Consider again case α: here a person is faced with a choice between strong reasons of self-interest and strong altruistic reasons: persons *e* and *h* will act on self-interest, while agent *a* will act on altruistic reasons. Notice that all three decision rules satisfy

Sterba's requirement for a reasonable commensuration: the highest reasons on one dimension are chosen over the lowest on the other.

Following Sterba, we might think of a morality as specifying a decision or commensuration function as in Figure 1, demanding when we must act altruistically. Thus understood moral philosophy is, ultimately, interested in seeking to justify some decision function, or at least some family of functions. But before analyzing the commensuration that is demanded of us, I propose to first consider what functions seem plausible for communities of humans. As Sterba rightly indicates (56-7) like most other traits, a tendency to act altruistically varies among us. With no pun intended, let us suppose it to be distributed on a Gaussian curve: there will always be those on the tails who are much more altruistic, or much more egoistic, than morality assumes. Human morality, though, is something directed at the typical or normal moral person. To slightly alter Kant's insight, our morality is directed neither to the saints nor the villains among us, but to a normally good person going about her life. Given this, a morality cannot demand much more altruism than, given her nature and character, our normal moral agent can deliver. As Rawls would say, the "strains of commitment" cannot be overwhelming.[4] A free and equal person cannot endorse a morality that she cannot live up to, or one that she could honor only with great difficulty. So before we know where we should, as it were, draw the line, we must have some ideas of the types of lines most humans can live with.

3.2 Hamilton's Rule

The most elegant, and in many ways accurate, analysis of the strength of our tendency to act altruistically was advanced by W.D. Hamilton.[5] What is now known as "Hamilton's rule" (eq. 1) essentially specifies a function for the strength of altruism, where r is the coefficient of relatedness between the helper and helped (essentially the probability that the helper shares the helped's genes), b is the benefit to the helped and c is the cost to the helper. Altruism is favored by evolution when the weighted benefits exceed the costs, as specified in

$$\text{Eq 1.} \quad rb>c$$

In humans, the value of *r* is 1 for identical twins, .5 for parent-siblings and between siblings, and (without inbreeding) .25 between grandparents and grandchildren, and between half-siblings. It is .125 between first cousins. Hamilton's rule quantifies the extent to which altruism — one organism sacrificing its fitness for another — can determine the behavior of an organism without endangering its genotype.

Hamilton's rule is a powerful (though, as we shall see, not the only or even the most adequate) explanation of a fundamental yet in many ways puzzling aspect of the development of an altruistic species: at least on the face of it, an organism that sacrifices its own fitness for the sake of others should go to extinction. The Social Darwinists were wrong about much, but they appreciated the problem that natural selection poses for altruism (though they were deeply wrong about the solution): that humans are altruistic at all is a puzzle that requires explanation. Take any group divided between altruists and selfish agents, where the altruists have a generalized tendency to help their fellow group members at some fitness cost to themselves, while the selfish individuals do not. Under such conditions the altruists will be good for the group, but in every generation they will constitute a diminishing percentage of the population. Unless we tell a more complicated story, they go extinct.[6]

To see this better, consider the problem posed for even modestly cooperative behavior in Figure 2:

		Column	
		Cooperate	*Defect*
Row	*Cooperate*	3 3	4 1
	Defect	1 4	2 2

Figure 2

This, of course, is a Prisoner's Dilemma (4=best, 1=worst for each). Assume that these payoffs correspond to the fitness of individuals in an evolutionary setting. We can readily see that a society of mildly altruistic people, who cooperate with others, will receive on average payoff of 3. (I call this "mildly" altruistic because both do better at the cooperate/cooperate payoff than the defect/defect outcome.) The problem is that our society of cooperators can always be invaded by a mutant defector: the average payoff will be 3 for the general population, but 4 for the mutant, allowing it to grow. Thus pure cooperation is not an evolutionarily stable strategy.[7] On the other hand, pure defection *is* a stable equilibrium: a society of pure defectors has an average fitness of 2, an invading cooperator would have a fitness of 1, and so could not get a foothold in the population.

Hamilton's rule explains one (let us say, a special) case under which altruistic behavior will not go extinct in a species: when it is directed at close kin. And Hamilton's rule explains a good bit of human altruism. There is a very strong tendency of parents to sacrifice for their children, and siblings to sacrifice for each other (in the United States, 86 percent of kidney transplants go to close kin, while less than .5 percent come from anonymous strangers).[8] Hamilton's rule also explains large-scale social cooperation among the eusocial insects, such as ants, bees, and wasps, which are haplodiploid.[9] In such insect groups sisters have an r of .75, thus making altruism, as it were, the normal course of things. But the core problem for the study of humans is that we are an ultra-social species who often rely on each other, but in which genetic relatedness does not account for altruism between strangers. In my view — and I think this would be the view of almost every scholar who has studied the evolution of altruism — Sterba is altogether too sanguine in his assumption that we are a highly altruistic species. He asks: "But is there really this difference in motivational capacity? Do human beings really have a greater capacity for self-interested behavior than for altruistic behavior?" The answer, I believe, is a clear and resounding "yes" ("no" for the eusocial insects). Humans are ultra-social creatures who do indeed help those who are not closely genetically related, but this helping is limited. In the evolution of human beings, there were always two opposed forces: a selection towards self-interested behavior and one towards cooperative, altruistic, behavior.

3.3 Direct Reciprocity

The development of human ultra-sociality has, as it were, depended on getting the most, and most socially effective, altruism at the least cost to altruistic individuals. We are typically interacting with those whose *r* approaches 0, not .75. Given this, the study of the development of human altruism is complex and controversial. There is, I think, consensus that no single mechanism explains human altruism towards strangers, though there is great debate about the operative mechanisms (genetic group selection, cultural group selection, direct reciprocity, indirect reciprocity) and their relative importance.[10]

For the last thirty years, those working on the evolution of cooperation, inspired by the path-breaking work of Robert Axelrod, have put a great deal of weight on the idea of direct reciprocity. As is well known, Axelrod demonstrated that in repeated Prisoner's Dilemmas the strategy of tit-for-tat (cooperate when you first meet another, and then do to her on the i^{th} play whatever she did to you on the i-1 play) allows cooperation to evolve in a large variety of settings.[11] More generally, assortative interactions, in which cooperators tend to cooperate only with other cooperators, is effective in allowing cooperation to evolve.[12]

Indeed, this idea of "direct reciprocity" — helping a person if she helps you — is actually a version of Hamilton's rule, as expressed in equation 2:[13]

Eq. 2. (Modified Hamilton Rule for Reciprocity) $eb>c$

In equation 2 e is the expectation that the helped will return the favor, which replaces Hamiton's r, the degree of genetic relatedness (b continues to be benefits to the person helped, while c designates the the costs to the helper). According to equation 2, Alf will engage in altruistic action ϕ toward Betty if the benefits of the act to Betty, weighed by e, exceed the costs to him. On the face of it, this seems a promising general analysis of how altruistic behavior can involve in groups that are not closely genetically related.

It seems doubtful, however, that direct reciprocity can be the prime basis for the evolution of cooperation in medium-to-large groups.[14] And in any event, while much of human altruism has strong features of reciprocity, morality in particular appears to require unrequited altruism.[15] But unrequited altruism is the most perplexing of all forms of altruistic action.

3.4 The Evolution of Unrequited Altruism

The core question for the evolution of human pro-social tendencies is how are we to explain unrequited altruism. As we saw in Figure 2, our unconditional cooperators are always at a fitness disadvantage, and so always can be successfully invaded by defectors. Unrequited altruists (unconditional cooperators) are always at a fitness disadvantage of 1. But now suppose another sort of cooperator enters into the population: a *punishing cooperator*.[16] A punishing cooperator cooperates, but also punishes defectors. Let us suppose that the punishers inflict a punishment that is sufficient to more than remove the benefits of defection. In this case, the cooperators can drive the defectors from the population, and cooperation can stabilize and resist invasion by defectors. The problem, though, is that a society of our punishing cooperators can be invaded by "easy-going" cooperators who cooperate (and so get all the gains from cooperation) but never inflict punishment. So it now looks as if the cooperating punishers will simply be driven from the population by the easy-going cooperators — and should the easy-going cooperators take over, *they* can be successfully invaded by the defectors! However, as Figure 3 shows, the gap in fitness between the cooperating punishers and the easy-going cooperators is not constant: as the number of defectors in the population decreases because of the altruistic work of the punishers, the cooperating punishers punish less often. But the less they punish, the more their fitness converges with the easy-going cooperators (they only suffered a fitness differential because of the costs of punishment). Indeed, the gap in the fitness between easy-going cooperators and cooperating punishers reduces to zero if defectors are eliminated from the population. Of course we would not expect this to actually reach zero — mutations, errors in determining defection, and immigration from other groups will result in greater than zero rates of punishing. Nevertheless, under a large range of values of the relevant variables, the discrepancy in relative fitness may be small enough so that punishers remain a high proportion of the population and defection thus declines.[17] However, it is certainly possible that a group may evolve to a "mixed" or polymorphic equilibrium, composed of punishers, easy-going cooperators, and perhaps defectors as well.[18]

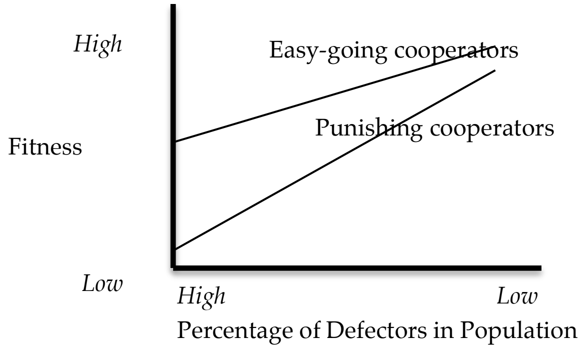

Figure 3

Punishes are *genuinely altruistic* in the sense that they engage in activities that do not best maximize their own goals, but are advantageous for the group. Punishing a violator of cooperative norms is a public good; as the easy-going cooperators share in the good, they do better by cooperating but never punishing. And of course an agent solely devoted to her own interest will defect if she is not punished; cooperators will not. Empirical investigation provides support for the importance of altruistic punishment in cooperative interactions. A study conducted by Falk, Fehr, and Fischbacher shows the importance of punishment for norm violation.[19] In this study in the first round subjects played a Prisoner's Dilemma, which was depicted as contributions to a cooperative enterprise with the possibility of free-riding. In the second round subjects were informed of others' choices (identities of participants remained anonymous), and were given the opportunity of punishing others for their play on the previous round. In the low punishment treatments, subjects could spend up to *y* unit of money to punish the defectors up to a *y* amount (a payment of *y* yielded a punishment of *y*); in the high punishment treatments *y* unit spent in punishment could reduce another's gains by up to 2.5*y* and 3*y*. Here one can expend money to reduce others gains by a multiplier. Falk, Fehr, and Fischbacher found:

- When cooperators punished, they only punished defectors. In both treatments, 60-70% of the cooperators punish.
- In the high punishment treatment, defectors also chose to "punish." When they did so, they punished other defectors and cooperators randomly. We might call this "spiteful punishment." Such defectors thus will *punish if doing so betters their relative position vis a vis the punished*.
- Cooperators by far imposed higher punishments.
- In low punishment treatments (where a unit spent punishing could not reduce the relative payoffs of the punishers and the punished), only cooperators punished.

After reviewing their data, Falk, Fehr, and Fischbacher conclude that "theories of inequality aversion cannot explain why cooperators punish in the low sanction condition. Despite this, 59.6 percent of the cooperators punished the defectors…. This suggests that the desire to retaliate, instead of the motive of reducing unfair payoff inequalities, seems to be the driving force of these sanctions."[20]

3.5 From First- to Second-Level Altruism

It is natural to think that human altruism must basically consist in the tendency to be altruistic in one's relations with others: for example, playing cooperatively in Prisoner's Dilemmas (Figure 2), even when one could get away with defecting. But we have seen that society of easy-going cooperators would be vulnerable to invasion by ruthless defectors. Students of the evolution of altruism increasingly have come to understand altruism not as a first-order tendency to cooperate, but as a tendency to enforce cooperative social norms. Indeed, what is really crucial is that punishment renders first-order cooperation no longer necessarily an altruistic act; the altruistic component of human society is thus focused on the punishing of those who break the cooperative norms. Note that in this case unrequited altruism — altruistic punishment — is the core of the account.

Of course we not go so far as to say that first-level altruism — the tendency to cooperate rather than defect on cooperative arrangements — is unimportant to human social cooperation. Surely our altruism is not just expressed in our tendency to punish those who violate social rules, but in our tendency to voluntarily follow these rules ourselves, even when we could escape punishment. There is good reason to think that the basis of human sociality is the altruism of *rule-following punishers*: those who follow the rules and punish those who do not.[21] The power of such altruism is that it allows us to avoid the cooperative dilemmas expressed in Figure 2, while minimizing the costs of altruistic action — the decrease in the fitness of altruists as compared to non-altruistic members of the group.

4 Altruism and the Moral Order

4.1 The Asymmetry of Morality

I have been concerned with how we might explain and understand a unique feature of human life: ultra-sociality and altruism under conditions of low genetic relatedness. Now I still have not said how much altruism morality should require of us. But I believe there is an important, if very general, implication of the analysis. We are right to think of altruism towards strangers as the basis of human ultra-sociality, especially in our complex modern world,[22] but there are compelling reasons to think that such altruism is a precious resource that must be efficiently and effectively employed. It is because altruism is such a precious and relatively scarce resource for humans that morality *demands* of us a *minimum* of altruism, and almost always leaves the pursuit of self-interest up to the actor. Elliot Sober and David Sloan Wilson, who conceive of morality in this way, point out that "Commonsense morality seems to set minimum standards concerning how much self-sacrifice is required, but it allows individuals to sacrifice *more* if they wish."[23] If we adopt Sterba's point of view and conceive of morality as simply demanding a non-arbitrary balancing of both reasons, we are left with the puzzle: "Why doesn't morality place a lower bound on how much *selfishness* we are required to exhibit, but allow people to be more selfish if they wish?"[24] Once one adopts the view of morality as reconciling egoism and altruism, we are immediately confronted with the puzzle of morality's asymmetric attitude: it characteristically *demands* minimum altruism but allows more but very seldom (on some views, never) demands minimum egoism.

At least on my reading of his essay, Sterba does not fully appreciate this asymmetry. He holds that "*a certain amount of self-regard is morally required*, and sometimes, if not morally required, at least morally acceptable. Where this is the case, high-ranking self-interested reasons have priority over conflicting low-ranking altruistic reasons, other things being equal" (53, emphasis added). Some moral theories accord an important place for duties to self; many others deny that there are any such duties. In general, they are not the core of moral systems. Apparently Sterba does not think that it is important whether morality requires or merely permits one to act on self-interest, but surely it matters a great deal that it does not merely *permit* us to act altruistically: it demands. And it is an unusual *moral* complaint to insist that another just needs to be more egoistic (though it is common enough in *self*-help books). Indeed, Sterba too recognizes that first and foremost morality demands minimum altruism (53); still, he is so insistent that the crux of morality is to dictate the correct commensuration function that he loses sight that its main concern is to require a minimum of altruism in all acceptable commensurations (see 71, note 26). The important question is how much altruism we can demand of others; we can leave it to each to look out for her self-interest. Perhaps curve *h* in Figure 1 is the most we can hope for in a humanly realistic morality (an eusocial morality would no doubt be in the vicinity of *a*). Perhaps even *h* is too altruistic.

4.2 Cooperative Norms and Human Altruism

We cannot commence a realistic social and political philosophy with the supposition that unrequited altruism to strangers comes easy to us, or that we easily see reason to override our self-interest for the good of the anonymous other. Our altruism is expressed in the fact that we are rule-following punishers: we generally comply with cooperative norms, though almost all cooperative norms require punishment to stabilize them. Moral rules that overstrain our altruistic resources will be unstable. Of course philosophers and priests can preach that much greater altruism is required, but there is good evidence that mere exhortation makes little difference: people act as they think others act.[25] If this is right, the main ways that political orders build on altruism is to ensure that fair and socially beneficial rules are followed by all with a genuinely altruistic concern that everyone plays according to the fair rules.

Do modern welfare states overstretch these resources? Perhaps some do, but it is clearly wrong to say that unrequited altruism is alien to humans: we are most definitely creatures of limited altruism. But we should not forget Rawls' lesson that the core idea of a just society is that it is a truly cooperative social and economic order characterized by reciprocity. Rawls was entirely right that social orders thus understood — as opposed to social orders in which many conceive of a large class of others as mere recipients of unrequited altruism — can be both free and stable. The debate within political philosophy — or, rather, that aspect of it focusing on distributive justice — properly concerns the conditions under which all participants can conceive of their order in this way. The crucial issue is under what conditions all reasonably benefit from the norms and institutions to which they are subject. No doubt in our very wealthy societies this will involve some sort of social minimum to help ensure (it can never be fully ensured) that no one subjected to the norms and institutions ends up in truly dire straits. Friends of the market believe, overall, that a market order with a modest state framework performs this task the best; those who favor a more robust state take a different view.[26]

I have no intention joining this debate here. But I must confess that I do not think we make much progress on resolving our deep perplexities by focusing on the sort of problem Sterba presents: a broadly specified case of a rich person with luxuries confronting a poor person with unmet needs. Sterba describes the case in the broadest of brush strokes: "the poor lack the resources to meet their basic needs to secure a decent life for themselves even though they have tried all the means available to them that libertarians regard as legitimate for acquiring such resources" (59). I worry about focusing on this sort of case. We need to know *a lot* more about the background institutions and norms, and the facts of the situation, which has given rise to this unacceptable state of affairs. Are the property rules unfair, discriminatory, or exploitative? Is the problem that the poor are not benefiting from an otherwise fair system of cooperation in which they participate? Why — is the educational system a culprit? Is the problem that there is deep lack of trust, and so widespread norm violation that greatly retards cooperative projects? Is the problem (perhaps with recent immigrants) that they have yet to be integrated into the cooperative enterprise, say because of residency restrictions, language barriers, or restraints on entering professions and trades? Is part of the problem, such as that in the United States today, that Draconian and oppressive drug laws have decimated the employment prospects of much of the male population and neighborhoods of a large group of citizens? At the outset of his essay Sterba aspires for philosophers to be taken more seriously in political decision-making (47). Until they are willing and competent to deal with this broad range of normative, political, sociological, and economic questions, it may be better for them to entertain more modest aspirations.

4.3 The Limits of Morality as Compromise

Thus brings me to a worry about Sterba's notion of "morality as compromise." To be sure, as I have argued it points to a deep insight. Humans are ultra-social creatures who cooperate in large groups not characterized by genetic relatedeness. Given this most fundamental of facts, it is enlightening indeed to see morality as a way in which the self-regard of such agents is channeled into cooperative action with an efficient use of altruism. However, we should not really think that morality is quite so simple as simply balancing egoism and altruism. Many of the deepest conflicts are *within* altruism. Consider Sterba's case of some who have as yet unmet basic needs advancing claims on those who are in the position to pursue luxuries, such as an excellent education for their children and an attractive home for their family. It is not just "selfishness," but human *altruism* that leads us to work for, and claim our right to keep, luxuries for our family and friends even in the face of the unmet needs of strangers. One spends sleepless nights worrying about one's children's welfare, and one undertakes a variety of onerous tasks with the aim of providing the resources that will allow them to achieve a good and interesting life. We must resist the temptation to think that all altruism is towards strangers, and all preference for our family is simply a form of selfishness. As Hamilton showed us, our altruistic tendencies vary with respect to those with whom we are interacting. We have some capacity for unrequited altruism towards strangers; more capacity for altruism towards those who have engaged in reciprocal relations with us, and much greater altruism towards close kin. We must be careful not to overtax the first of these by deeming it alone to be "truly altruistic." The main function of *demanding* unrequited altruism towards strangers is to ensure that all conform to the basic moral rules of cooperative social life. A social and political morality that too readily demands that we act altruistically toward strangers rather than act altruistically toward those whom we love is not, in my view, genuinely human.

NOTES

[1] *Proceedings and Addresses of the American Philosophical Association*, vol. 82 (2): 47-83. All parenthetical references in the text refer to pages of this essay.

[2] Sterba does not employ the notation I use in the text. For the moment I will accept that reasons can be simply ordered. See below, §2.

[3] For example, his invocation of the analogy with the principle of insufficient reason (54) suggests that in the absence of further information, the two sets of considerations should be treated as equally important. Cf. also his remark that "it may be objected that my argument for favoring morality over egoism and altruism would be analogous to naturalists and supernaturalists *splitting the difference between their views* and counting supernaturalist reasons *as valid half the time*, and naturalist reasons *as valid the other half the time*" (56, emphasis added). I try to show in this essay that any equal weighing of self-regarding and altruistic considerations would be deeply implausible.

[4] Rawls, *A Theory of Justice*, pp. 153ff.

[5] W.D. Hamilton, "The Genetical Evolution of Social Behaviour I," *Journal of Theoretical Biology*, vol. 7 (1964): 1-16. For a very helpful discussion, see Natalie Henrich and Joseph Henrich, *Why Humans Cooperate* (Oxford: Oxford University Press, 2007), pp. 45ff.

[6] See Sober and Wilson, *Unto Others*, pp. 18ff.

[7] According to one way of formalizing this idea, S is an evolutionarily stable strategy if and only if, with respect to a mutant strategy S* that might arise, either (1) the expected payoff of S against itself is higher than the expected payoff of the mutant S* against S *or* (2) while the expected payoff of S against itself is equal to the expected payoff of S* against S, the expected payoff of S against S* is higher than the expected payoff of S* against itself. The idea is this. Suppose that we have an S population in which one or a few S* types are introduced. Because of the predominance of S types, both S and S* will play most of their games against S. According to the first rule, if S does better against itself than S* does against S, S* will not get a foothold in the population. Suppose instead that S* does just as well against S as S does against itself. Then S* will begin to grow in the population, until there are enough S* so that both S and S* play against S* reasonably often. According to the second rule, once this happens, if S does better against S* than S* does against itself, S will again grow

at a more rapid rate. To say, then, that S is an ESS is to say that an invading strategy will, over time, do less well than will S. There are other ways of formulating the basic idea of an evolutionary stable strategy, but that need not detain us here.

[8] Henrich and Henrich, *Why Humans Cooperate*, p. 45.

[9] In which a female has two alleles but a male only one.

[10] For a very helpful survey see Henrich and Henrich, *Why Humans Cooperate*, chap 3; Sober and Wilson present the case for the importance of genetic group selection in *Unto Others*.

[11] See Robert Axelrod, *The Evolution of Cooperation (*New York: Basic Books, 1974).

[12] See, for example, Brian Skyrms, *The Evolution of the Social Contract* (Cambridge: Cambridge University Press, 1996), chaps. 3 and 4.

[13] See Henrich and Henrich, *Why Humans Cooperate*, p. 42.

[14] See Robert Boyd and Peter J. Richerson, "The Evolution of Reciprocity in Sizable Groups" in their *The Origin and Evolution of Cultures* (Oxford: Oxford University Press, 2005), chap. 8. See also Peter J. Richerson and Robert Boyd, *Not by Genes Alone*: *How Culture Transformed Human Evolution* (Chicago: University of Chicago Press, 2006), pp. 197ff.

[15] But cf. Skyrms, The Evolution of the Social Contract, pp. 61-62.

[16] I am following Robert Boyd, Herbert Gintis, Samuel Bowles and Peter J. Richerson, "The Evolution of Altruistic Punishment" in *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, edited by Herbert Gintis, Samuel Bowles, Robert Boyd and Ernst Fehr (Cambridge, MA: MIT Press, 2005): 215-227.

[17] See ibid.; Boyd and Richerson, "Why People Punish Defectors" in their *The Origin and Evolution of Cultures*, chap. 10.

[18] See Boyd and Richerson, "Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable groups" in their *The Origin and Evolution of Cultures*, chap. 9.

[19] Armin Falk, Ernst Fehr, and Urs Fischbacher, "Driving Forces Behind Informal Sanctions," IZA Discussion Paper No. 1635, (June 2005). Available at SSRN: http://ssrn.com/abstract=756366.

[20] Ibid., p. 15. This is not to deny that there is disagreement on the best interpretation of the data. For more egalitarian interpretations, see T. Dawes, Christopher, James H. Fowler, Tim Johnson, Richard McElreath, and Oleg Smirnov. "Egalitarian Motives in Humans," *Nature*, vol. 446 (12 April 2007): 794-96; James Fowler, Tim Johnson, and Oleg Smirnov, "Egalitarian Motive and Altruistic Punishment," *Nature*, vol. 433 (6 January 2004): E1.

[21] I develop this idea in *The Order of Public Reason: a Theory of Freedom and Morality in a Diverse and Bounded World* (Cambridge: Cambridge University Press, 2011), chap. III.

[22] Evidence supports the hypothesis that norms about fairness to strangers develop along with markets. Ibid., chap. VIII.

[23] Elliot Sober and David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge: Harvard University Press, 1998), p. 240. Emphasis in original.

[24] Ibid.

[25] See Cristina Bicchieri and Erte Xiao, "Do the Right Thing: But Only if Others Do So." *Journal of Behavioral Decision Making*, vol. 22 (2009): 191-208. . In their experimental work on public goods games among the Machiguenga and the Mapuche, Joseph Henrich and Natalie Smith also found that "the primary indicator of what a subject will do is what the subject thinks the rest of the group will do." "Comparative Evidence from Machi-guenga, Mapuche, and American Populations" in *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, edited by J. Hennrich, R. Boyd, S. Bowles, et al. (Oxford: Oxford University Press, 2004): 125–67 at p. 153.

[26] Perhaps there are certain wild-eyed libertarians who, regardless of the background institutions and the norms in play will always respond "Let them eat cake!" (though we should remember that Marie Antoinette does not have a place in the libertarian pantheon). If some libertarians think that claims to property need not be justified in terms of norms, rules, and institutions that give reasonable benefits to all participants, and which all as free and equal moral persons can endorse, then Sterba is entirely right to dismiss such views. But, as I have been stressing, the serious and difficult issue is what background norms and institutions satisfy these conditions.