Respect for Persons and the Evolution of Morality*

Gerald F. Gaus

1. Impartial Reason or/and The Evolution of Morality?

Let me begin with a stylized contrast between two ways of thinking about morality. On the one hand, morality can be understood as the dictate of, or uncovered by, impartial reason. That which is (truly) moral must be capable of being verified by everyone's reasoning from a suitably impartial perspective. If we are to respect the free and equal nature of each person, each must (in some sense) rationally validate the requirements of morality. If we take this view, the genuine requirements of morality are a matter of rational reflection and self-imposed law. For Kant it seemed to be a matter of reflection by a rational individual, testing the impartiality of his maxims. For Rousseau, who was an important influence on Kant, under the proper conditions collective deliberation could yield impartial rules of justice that are willed by all.

From another point of view moralities are social facts with histories. The heroes of this tradition are Hume, Ferguson and, perhaps surprisingly given his "deductive" method, Hobbes. The moral codes — or if "code" implies too much systematization, moral "practices" — we have ended up with are, to some extent, a matter of chance. This is by no means to say that morality is entirely arbitrary, but it does contain a significant arbitrary element. The morality we have ended up with is *path-dependent*: only because our moral codes have started somewhere, and have changed in response to unanticipated events, can we explain why we ended up where we have, and different societies end up in different places.

The proponents of each view typically seek to discredit the other. Those who conceive of morality as the demand of impartial reason often insist the evolutionists confuse "positive morality" (the moral code that people actually follow) with justified (or true) morality, which is revealed by impartial reason. The positive morality that has evolved is simply what people *think* is morality, not what *really is* morality. Less sophisticatedly, some simply insist that the evolutionary view exemplifies (the horrible vice of) "undergraduate cultural relativism," and so dismiss the whole idea.

---

*I have presented my thoughts on these matters to a number of audiences, and have greatly benefited from the questions and comments of the participants. I would like the thank the participants in the Research Triangle Ethics Circle, the Committee on Law and Philosophy at Arizona State University, the Institute of Humane Studies Current Research Seminar, and the Philosophy Departments at the University of Reading, North Carolina State University, the University of Georgia, and especially my colleagues and students at the University of Arizona. My special thanks to my good friend Fred D'Agostino for encouraging this line of inquiry.

Impatience with the opposing view is not restricted to advocates of impartial reason. Ken Binmore, who has recently defended an evolutionary account of morality announces at the outset that "orthodox moral philosophy has gotten us nowhere because it asks the wrong questions. If morality evolved along with the human race, asking how we ought to live makes as much sense as asking what animals ought to exist, or which language we ought to speak."[1] Binmore cannot mention "Kant" without vitriol: Kant engages in "Humpty-Dumpty" and "Alice-in-Wonderland reasoning"; he is "an emperor clothed in nothing more than the obscurity of his own prose."[2] "Can it really be," Binmore asks, "that his claim to fame as a moral philosopher is based merely on his having invented one of the fallacies of the Prisoner's Dilemma [i.e., the universalization principle] before anyone else?"[3] Only the "scientific tradition" of Hobbes, Hume and Smith can save us.[4] And Binmore is by no mean alone; Russell Hardin's views of Kant are only marginally more charitable.[5]

In this paper I argue that Kantian-inspired conceptions of morality — or, as I shall call them, "public reason" conceptions — must embrace significant parts of the evolutionary view. Morality is properly seen as consisting of self-imposed requirements verified from the impartial perspective *and* as having a history that is path-dependent. Indeed, I argue that only an evolved morality can be justified to everyone, and so only an evolved morality provides the basis for each treating all as free and equal moral persons.

I begin in Section 2 by sketching a family of moral views that are committed to what I call the Public Justification Principle. It is important to begin by reminding ourselves why respect for others requires the public justification of moral requirements from the impartial perspective, and why only moral requirements that in some sense are universally self-legislated are consistent with treating our fellows as free and equal moral persons. Having sketched the grounding of the Public Justification Principle, Section 3 then considers what seems to be an insuperable problem for public reason views of morality: reasonable persons are characterized by a deep pluralism about the basis for self-legislation. To show just how serious the problem is, Section 4 reviews two great, unsuccessful, attempts to solve the problem of public justification given deep evaluative pluralism: those of Rousseau and Rawls. Section 5 points the way to a more adequate approach to the problem, but we shall see that the solution is

---

[1] Ken Binmore, *Natural Justice* (Oxford: Oxford University Press, 2005), p. 1.

[2] Ibid., pp. vii, 37-38.

[3] Ibid., p. viii.

[4] Ibid., p. 39.

[5] Russell Hardin, *Indeterminacy and Society* (Princeton: Princeton University Press, 2003), ch. 6.

indeterminate; Section 6 argues that social evolutionary processes can complete the justification process. I reflect on some of the implications of the analysis in Section 7.

2. Respect for Persons and Public Justification

2.1 Morality, Authority and the Threat of Subjugation

Social morality provides a set of principles that allows one person to make moral demands on others. As John Stuart Mill rightly recognized, when one appeals to social morality one makes a claim to something like moral authority over another:[6] one is claiming that on this matter, the other is not to do as she wishes, but as you require. Stephen Darwall has recently stressed the way in which interpersonal morality involves "authority relations that an addresser takes to hold between him and his addressee."[7] To make a moral demand on another is to assume a practical authority over another to make demands and to demand compliance.[8] To make a moral demand is not simply to call attention to your claim and its merits, but to insist that the claim is backed up with an authoritative moral reason for the other to do as you demand.[9] Now although this form of authority is as commonplace as our moral life, it is by no means unproblematic. One person (Alf) is supposing that his view of what the other (Betty) *must* do (whether Betty wishes to or not) trumps her view of her reasons to act, and what she should do. If she does not, he will normally deem her blameworthy, and liable to moral criticism. As Darwall points out, when Alf makes a moral claim on Betty he is not requesting or calling attention to his claim: he is demanding that Betty complies. Alf thus seems to be claiming that Betty is subject to his authoritative demands. She must obey even when she disagrees. But now we are faced with the question: by what right does Alf claim such authority over the life of Betty?

Alf's answer to the challenge, no doubt, will be that it is not *his* authority, but the authority *of morality* to which Betty is subject. But "morality" only speaks through its interpreters, and Betty dissents from Alf's interpretation. As Hobbes recognized, "All laws, written and unwritten, have need of interpretation."[10] So the question becomes: on what grounds does Alf claim that his interpretation of the demands of morality has authority over Betty? Alf is claiming that his reason is "right reason" — but in almost

---

[6] See John Stuart Mill, *On Liberty* in *The Collected Works of John Stuart Mill*, J.M. Robson, ed. (Toronto: University of Toronto Press, 1977), vol. 18: ch. 1.

[7] Stephen Darwall, *The Second-person Standpoint: Morality, Respect and Accountability* (Cambridge, MA: Harvard University Press, 2006), p. 4.

[8] Ibid., pp. 10-11.

[9] Ibid., p. 76.

[10] Thomas Hobbes, *Leviathan,* Michael Oakeshott, ed. (Oxford: Basil Blackwell, 1948), p. 180 (ch. 26).

every dispute, each party claims that his or her reason is right reason. Hobbes was deeply worried about this problem:

> when men that think themselves wiser than all others clamour and demand right reason for judge, yet seek no more but that things should be determined by no other men's reason but their own, it is…intolerable in the society of men….For they do nothing else, that will have every of their passions, as it comes to bear sway in them, to be taken for right reason, and that in their own controversies: bewraying [sic] their want of right reason by the claim they lay to it.[11]

As always, Hobbes's concern is social stability — a concern that should not be dismissed or trivialized. His general point, though, is profound and goes beyond the concern with stability. Because o*f course* each party to a dispute claims that his reason is right reason, for Alf to demand that others conform to his reason *because* it is right reason betrays his lack of true reason by ignoring the nature of the dispute: the deep disagreement about the demands of right reason and the interpretation of social morality. For Kantians, however, not only is Alf's attitude anti-social and rationally suspect, it evinces a lack of respect for the moral freedom and equality of Betty. Alf appears to be claiming that he is a superior interpreter of morality, and so Betty is under his moral authority, though the crux of their dispute is precisely about who is the superior interpreter. Although it is something of a rhetorical overstatement, we can appreciate the force Jeffrey Reiman's worry that Alf's assertion that he "has a higher authority" over how Betty should act raises the specter of "subjugation" — that "the very project of trying to get our fellows to act morally" may be "just pushing people around."[12]

This worry about using claims to superior moral insight as a way of "pushing others around" is, I think, quintessentially liberal. Recall that Locke's canonical liberal text, *The Second Treatise*, with its adamant denial of natural authority, was written as a response to Robert Filmer's assertion that some were naturally the moral superiors of others. Filmer vigorously upheld his view against those who advocated the "dangerous opinion" of the "natural freedom of mankind."[13]

> Every man that is born, so far from being born free, that by his very birth he becomes a subject to him that begets him: under which subjection he is always to live, unless by immediate appointment from God, or by grant or death of his Father, he became possessed of that power to which he was subject.[14]

---

[11] Ibid., p. 26 (ch. 5).

[12] Jeffrey Reiman, *Justice and Modern Moral Philosophy* (New Haven, CT: Yale University Press, 1990), p. 1.

[13] Robert Filmer, *Patriarcha* in Peter Laslett, ed., *Patriarcha and Other Political Works* (Oxford: Blackwell, 1949), p. 53.

[14] Filmer, "Directions for Obedience to Government in Dangerous or Doubtful Times," in ibid., p. 231.

If there is any sense in saying that men are born free, Filmer insisted, it is that men are not born subjugated as servants, but as sons.[15] Filmer did not deny that fathers (and so monarchs) are bound by the (true) laws of nature to act justly towards their subjects and to care for their welfare, but he insisted that the authority to interpret this law resided in the father: the upshot is that the family is governed by the reason of the father.[16] Though Filmer was distinctive in deriving natural moral authority from patriarchal authority, he is by no means unique in upholding a claim that some people have intrinsic moral authority over others. Aristotle's account of the status of slaves as "living tools" incapable of friendship,[17] Mill's own acceptance of authoritarianism for "races" in their "nonage,"[18] and even, I think, Sidgwick's principle that "enlightened Utilitarians" may advocate an "esoteric morality" that is the criterion of genuine moral requirements but is not revealed to *hoi polloi*[19] — all seems to conform to the picture of claims to superior insight into morality as being ways that some people employ to push others around.[20]

2.2 Universal Self-legislation

Social morality presupposes that we claim authority over others, yet liberals insist that we are all free and equal moral persons, and so each has an equal status as moral interpreter; each should be free to interpret her own moral obligations for herself. How can liberalism's commitments to moral freedom be reconciled with the authoritative nature of moral demands? Kant's ideal of the realm of ends provides the core insight.

> A rational being belongs to the realm of ends as a member when he gives universal laws in it while also himself a subject to these laws. He belongs to it sovereign when he, as legislating, is subject to the will of no other.[21]

Contrast to Rousseau:

> As for the associates, they collectively assume the name *people*, and individually call themselves *Citizens* as participants in the sovereign authority, and *Subjects* as subjected to the laws of the State….

---

[15] Filmer, *Patriarcha*, pp, 73-74.

[16] Ibid. p. 96.

[17] Aristotle, *Nicomechean Ethics*, Sir David Ross, trans. (Oxford: Oxford University Press, 1954), p. 212 [1161a30-b19].

[18] Mill, *On Liberty*, ch. 1, para. 10.

[19] Henry Sidgwick, *The Methods of Ethics*, 7th edn. (Chicago: University of Chicago Press, 1962), pp. 489ff.

[20] For a general characterization of moral authoritarianism, see my *Social Philosophy*, pp. 6ff.

[21] Immanuel Kant, *Foundations of the Metaphysics of Morals*, Lewis White Beck, ed., trans. (Indianapolis: Bobbs-Merrill, 1959), p. 52.

> This formula shows that the act of association involves a reciprocal engagement between the public and private individuals, and that each individual, by contracting, so to speak with himself, finds himself engaged in a two-fold relation: namely as a member of the Sovereign toward private individuals, and as a member of the State toward the Sovereign.[22]

Kant insists that, for morality to be consistent with "the dignity of a rational being" a rational being must obey no law other than that he gives himself. The individual is both legislator and subject. Rousseau's point is essentially the same: to avoid the degradation slavery, one must, qua member of the Sovereign, give himself the law that, qua member of the State, he must follow.

Kant's depiction of the self-legislative nature of a free morality stresses that each rational being has a will that is legislative for every other will, giving laws to all to which he is, qua subject, also subject. Our moral freedom consists in being "a legislative member in the realm of ends,"[23] but we are also subject to such legislation. Now it is important that by "realm" Kant meant "the systematic union of different rational beings through *common laws*."[24] So Kant does not think it is fine if you legislate in one way and I in another. Implicit in Kant's analysis of morality, then, is a unanimity requirement: we legislate common laws. The same morality thus must be legislated by all rational beings. In Rousseau's work the collective nature of the problem rises to the fore. For Rousseau, we all must, together, legislate for each of us, yet each, "nevertheless obey only himself."[25]

2.3 The Generic Public Justification Principle

If we combine Kant's requirement of universal self-legislation of rational beings with Rousseau's characterization of the problem in terms of collective deliberation, we are led to something along the lines of:

> *The (Generic) Public Justification Principle*: *M* is a (*bona fide)* moral requirement only if each and every member of the public *P*, under conditions *C*, has sufficient reason(s) *R* to accept *M* as a binding requirement on all.

---

[22] Jean-Jacques Rousseau, *On The Social Contract* in *The Social Contract and Other Later Political Writings,* Victor Gourevitch, ed., trans. (Cambridge: Cambridge University Press, 1997), p. 49.

[23] Kant, *Groundwork*, Akademie, pp. 433-34.

[24] Ibid., pp. 433-34.

[25] Rousseau, *The Social Contract*, p. 50.

The Public Justification Principle, as Rawls puts it, conceives of justified moral principles as mutually acknowledged "by free persons who have no authority over one another."[26]

Because I am concerned with a family of public reason views, I focus on a generic formulation of the principle. Because this is a generic principle, I leave open the crucial problem of just how to specify *P* (whether the members must all be reasonable, fully rational, etc.). The Public Justification Principle supposes that there is some specification (and almost certainly some idealization) of the public such that if each member so described deliberated under some conditions (*C*), each would rationally endorse *M*.[27] One Kantian specification of *P* is the realm of rational beings; insofar as we act as members of *P* we act in accord with our status as rational moral beings. In contrast, Rousseau often seems to conceive of *P* as an actual collective "moral person" that makes actual collective decisions (under very demanding conditions, *C*).[28]

For simplicity sake, in this essay I suppose that members of *P* under *C* are conceived of as deliberating about specific moral requirements. We can think of the problem posed to members of *P* under *C* as: what should be the moral requirement, *M*, regulating matter *X*? This is closest to the Kantian-inspired view of the problem as legislating. It is more accurate, however, to suppose, as Rawls did in "Justice as Fairness," that the object of justification is a moral practice: an interlocking set of moral requirements, permissions and prohibitions that distinguishes certain roles and obligations. Thus the members of the public should probably be thought of as considering sets of moral rules such as those that comprise the practices of ownership, personal privacy, protection of the person, and so on. Everything said here can be translated into the notion of a moral practice. What concerns members of *P* under *C* is whether they have reason to endorse the same requirements or practices.

---

[26]John Rawls, "Justice as Fairness" in Samuel Freeman, ed., *John Rawls: Collected Papers* (Cambridge, MA: Harvard University Press, 1999), p. 55.

[27] It might be argued that an egoist has reason to accept *M* as a binding requirement on all, but to ignore *M*. We must recall that we are considering certain idealized persons (e.g. reasonable); in a fuller account we would also have to explicate what is involved in "accepting" a moral requirement, and whether the egoist we are considering can be said to have accepted *M*. I am indebted to Jim Sterba for pressing me on these points and pointing out to me the inadequacy of an earlier formulation.

[28] Rousseau's conditions include, for example, that the members of the public are not organized into factions or parties and they have independently arrived at their views. As Rousseau says: "If we take the term in its strict sense there never has been a real democracy, and there never will be....How many conditions that are difficult to unite does such a government presuppose!" *The Social Contract*, Book III, ch. iv.

2.4 The Companion Deliberative Model

One of Rawls's fundamental insights was that the justificatory problem — what moral requirements do *P* under *C* have reason to endorse? — can be translated into a deliberative problem.[29] Suppose we understand a member *i* of *P* under *C* as consulting his relevant evaluative standards — the full set of considerations that is relevant to his decision whether to endorse some moral requirement (§3). After consulting his evaluative standards, *i* proposes his preferred moral requirement, M_i: the moral requirement that, on his (somewhat idealized) reasoning, best conforms to his evaluative standards. (This procedure is akin to that utilized by Rawls in "Justice as Fairness.")[30] Suppose also that, on the basis of his own evaluative standards, each *P* under *C* ranks everyone's proposed requirement.

This simple statement of the deliberative problem — as I said, inspired by Rawls's first formulation of his own theory — has real advantages over more familiar formulations. One of the problems with much contemporary contractualism is that it typically employs a notion of reasonable acceptability (or rejectability) without being clear about the feasible set: to ask what one can reasonably accept (or reject) without knowing the feasible alternatives is an ill-formed choice problem. "Rationally rejectable in relation to what options?" is the crucial question. In our deliberative problem the feasible set is defined by the set of all proposals. Rawls never made this common mistake: the parties to his original position in *A Theory of Justice* choose among a small set of traditional proposals, so their choice problem was well-defined. However, Rawls built into his later and more famous formulations of the deliberative problem a host of controversial conditions (as we will see in section 4, the aim to make the choice problem determinate must resort of such demanding and controversial conditions). Instead, our deliberative problem is a straightforward articulation of the Public Justification Principle which it is meant to model: if one accepts the Public Justification Principle as posing the correct justificatory problem, there is strong — indeed, I think compelling — reason to accept this deliberative model. The only element it adds is the interpretation of what one has a reason to accept in terms of a ranking of the proposals advanced by each member of *P* under *C*, translating the idea of "rational acceptance" into each person's ordinal rankings based on his evaluative standards. As I said, doing so is a compelling way to make the deliberative problem well-formed, providing a non-arbitrary feasible set from which the members of *P* under *C* are to

---

[29] John Rawls, *A Theory of Justice*, revised edn. (Cambridge, MA: Belknap Press of Harvard University Press, 1999), p. 16 (p.17 of the original edition).

[30] "Their procedure … is to let each person propose principles …." ("Justice as Fairness," p. 53.) As will be seen, in a number of ways I am proposing going back to the project begun in that classic essay, which posed a simple and compelling Kantian deliberative problem.

choose. But this leads directly to the really basic question: what are their evaluative standards?

3. Evaluative Pluralism and Moral Disagreement

As stated, most moral theories can endorse the Public Justification Principle and its companion deliberative model: if the parties are so specified that they all accept, say, a certain substantive moral theory, moral requirements justified by that moral theory would also be justified by the Public Justification Principle. The Public Justification Principle and its companion deliberative model would do little or no work. The Public Justification Principle becomes a substantive test of a moral requirement if we accept Rawls's claim that a wide range of rational disagreement is the "normal result of the exercise of human reason."[31] Suppose, then, that we accept reasonable pluralism in the sense that our characterization of the members of *P* deliberating under conditions *C* includes that members of *P* reason on the basis of different values, ends, goals, etc. This does not prejudge whether values are "ultimately" plural, for perhaps fully rational, omniscient beings would agree on what is valuable: the important point for public reason views is that the characterization of *P* under *C* allows for diversity in the basis of their reasoning about what moral requirements to endorse. Abstracting from the notions of goods, values, moral "intuitions" and so on, let us say that Σ is an evaluative standard for Alf if holding Σ (along with various beliefs about the world) gives Alf a reason to endorse *M*.[32] Evaluative standards, then, are to be distinguished from justified moral requirements: as I have characterized them they need not meet the test of Public Justification, but are the reasons for members of *P* to endorse (or not endorse) *M*. In the deliberative model evaluative standards are the bases for ranking proposals.

We suppose, then, plurality of evaluative standards for *P* under *C*. But how great is this pluralism? Under *radical* pluralism we would so characterize the deliberations of *P* under *C* as to allow for just about any evaluative standard that rational agents have endorsed — including those that value the suffering of others and subjugating others. But it is unlikely that a plausible conception of parties who conceive of each other as free and equal moral persons would endorse such oppressive standards. Remember, our core problem is that Alf and Betty confront each other as advancing different interpretations of their *bona fide* moral requirements; unless they can understand

---

[31] Rawls adds: "within the framework of free institutions of a constitutional regime." *Political Liberalism*, paperback edn. (New York: Columbia University Press, 1996), p. xviii.

[32] I leave aside here whether Σ is itself a belief about the world, or supervenes on one, as ethical naturalists would have it. Nothing in the analysis precludes moral realism as a metaethical or metaphysical thesis. The rationality-based constraint on justificatory reasons is the crucial principle on which the analysis rests.

each other as actually reasoning in a way that is relevant to morality, they do not have reason to conceive of their dispute as one *within* morality. So if Alf is arguing simply on the basis of narrow self-interest, or on the goodness of seeing others suffer, Betty will not see herself as confronting another moral interpreter, but facing someone denying the claims of morality. She must be able to understand their differences as within the bounds of moral disagreement. Insofar as Alf and Betty both understand their dispute to be about morality, they must see each other as reasoning on the basis of considerations that are intelligible and relevant to the matter at hand. So, while liberals must hold that members of *P* under *C* will reason on the basis of different evaluative standards, they cannot embrace radical pluralism. As Isaiah Berlin has stressed, the range of plausible pluralism is limited by the "common human horizon."[33]

The members of *P* under *C,* then, must concur on a list of evaluative standards that are relevant to moral deliberation. We need not require that everyone actually affirms each standard, but all must be intelligible as a basis for deliberation. The important point is that they concur on the list of relevant standards but deeply disagree on their ordering. There is empirical evidence that our value disagreements are of this sort: people agree on what is valuable but have striking disagreements about what is more and less valuable. [34] Although the parties deeply disagree in their orderings of intelligible evaluative standards, it would be going too far to say that their orderings meet a condition of unrestricted domain, i.e., that the intelligible list of evaluative standards is ordered in all logically possible ways. Public reason liberals may plausibly maintain that everyone, say, holds that not killing innocents is more important that securing personal pleasures, or (*a la* Rawls) liberals might suppose that all members of *P* under *C* hold that achieving fair terms of cooperation is an important *diseratum*. Even Berlin thinks that the "common human horizon" allows some common ranking of values: we sometimes agree that some rankings are more "humane" while others are "indecent."[35] We should not insist that liberals adopt a version of pluralism more extreme than Berlin's. Although, then, we should not require liberal public reason views to suppose entirely unrestricted rankings of the set of intelligible evaluative standards, a compelling liberal account nevertheless must attribute great diversity to the evaluative standards that individuals endorse. So we

---

[33] See my *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project* (London: Sage, 2003), ch. 2.

[34] See Milton Rokeach, *The Nature of Human Values* (New York: The Free Press, 1973), p. 110; Milton Rokeach, "From Individual to Institutional Values," in his *Understanding Values* (London: Collier Macmillan, 1979), p. 208.

[35] See *Contemporary Theories of Liberalism*, pp. 43-46; see also Jonathan Riley, "Interpreting Berlin's Liberalism," *American Political Science Review,* vol. 95 (June 2001): 283-97.

suppose that members of the public deliberating under *C* are characterized by great, if not entirely unrestricted, evaluative pluralism. Call this admittedly imprecise degree of disagreement *deep evaluative pluralism*.

The problem for liberal public justification now is manifest. If the parties employ their evaluative standards to evaluate different proposed moral requirements, so long as their disagreements in evaluative standards are deep, these disagreements will seem to inevitably result in great disagreement in their rankings of candidates for moral requirements. If a member of the public P1 holds ranking $\Sigma_1 > \Sigma_2$ (read as "Σ_1 is ranked above Σ_2") while P2 maintains that $\Sigma_2 > \Sigma_1$, then if these are the only relevant standards, and, if the degree of justification of the moral requirements within a perspective is monotonic with the ranking of evaluative standards, P1 will hold $M_1 > M_2$, while P2 will rank the requirements $M_2 > M_1$. To be sure, the members of the public may display consensus on some basic moral requirements (as Berlin suggests, they may all see as wrong pushing pins into babies for fun), but given the depth of evaluative pluralism, and the importance of people's evaluative standards in their deliberations about what moral requirements they have most reason to accept, we would expect that great disagreement in evaluative rankings would result in great disagreements in the rankings of possible moral requirements. If the basis for judging moral requirements is diverse, so too will be the evaluations of moral requirements. Deep moral disagreement would seem the inevitable result of deep evaluative pluralism. The public reason liberal seems to have embraced incompatible requirements: justified morality requires agreement, but evaluative pluralism leads to disagreement. What's a liberal to do?[36]

4. Two Great Ideas: Rousseau and Rawls

4.1 Rousseau's Great Idea

Rousseau's work suggests a possible way out of the problem: he explicitly viewed the task of forming the "general will" as one of collective choice. Rousseau's theory of the general will is, of course, open to multiple interpretations,[37] but on one plausible version, he stressed that although members of the public (qua individuals) have diverse individual rankings and interests, we could, through the proper deliberative procedure, construct out of a these diverse individual wills a shared "social will" that would serve as the basis of self-legislation for all members of the public. Alternatively, we

---

[36] One way out of the problem — which I think is Kant's — is to bracket pluralism and suppose that we have the same basic human aims. I criticize this Kantian "solution" in "Recognized Rights as Devices of Public Reason" in Derrick Darby, ed. *The Rights Recognition Thesis,* forthcoming.

[37] See my "Does Democracy Reveal the Will of the People? Four Takes on Rousseau," A*ustralasian Journal of Philosophy,* vol. 75 (June 1997): 141-162.

might say that the aim is to create a shared social ranking of an option set out of a number of diverse individual rankings. We might call this a Social Morality Function. If we had some conclusively justified way to aggregate different orderings, then our differences in rankings of feasible requirements would not stand in the way of an impartially justified requirement that all members of the public would with good reason endorse.[38] As William Riker stressed, though, Arrow-type problems come to the fore here.[39] Kenneth Arrow, as we know, showed that for social choice over three or more options (i) when we do not restrict the permissible individual rankings; (ii) when we do not allow that any one individual's ranking is decisive over all pairs; (iii) if everyone holds that $Mx \succ My$, then the social ranking must hold $Mx \succ My$ and (iv) the ranking between any two alternatives does not vary according the presence or absence of a third option, then (v) there is no method of social aggregation that will always produce a complete transitive social ordering.

This famous result has been noted, and disputed, on a number of grounds. In my view, the reasoning behind it is much more powerful and is far more robust than is usually acknowledged by moral and political philosophers.[40] However, it may be objected that I have admitted that unrestricted domain in the orderings of evaluative standards is too strong an assumption (§3); if members of *P* under *C* agree in some of their rankings of evaluative standards, they may concur in some of their orderings of moral requirements. At least in its most straightforward form, unrestricted orderings of moral requirements would be a supposition needed to apply the theorem. So, it may be held, the assumption of deep evaluative pluralism is too weak to justify calling on Arrow's work. I think this is wrong. If we have deep evaluative diversity — even if not entirely unrestricted rankings — democratic pairwise choice over three or more options can still generate Arrovian cycles. Formally, whether or not cycles occur depends on whether members of *P* under *C* display considerable consensus on the dimensionality of the dispute: if members of the public, whatever their disagreements, arrange the options over the same dimensions (say, libertarian-egalitarian), then pairwise democratic choice will yield consistent results; if they disagree not only about the ordering, but the dimension that is relevant to choice (some see the options as arrayed over a left-right dimension, others over a conservative-progressive dimension, others over rights-consequentialist dimension), then we can expect significant

---

[38] See Fred D'Agostino, *Incommensurability and Commensuration*: *The Common Denominator* (Aldershot, UK: Ashgate, 2003), p. 100.

[39] I defend Riker's analysis in "Does Democracy Reveal the Will of the People?"

[40] I have argued for the importance and generalizability of the result in *On Philosophy, Politics and Economics* (Belmont, CA: Wadsworth, 2007) ch. 5. I especially stress that even if we are only concerned with identifying the best in the set, and not developing a complete social ordering, we face Arrow-like problems.

incoherence in the social choice. Rousseau, I think, was well aware of the general problem, even if not the formal representation. He stressed throughout *The Social Contract* that, for the majoritarian procedure to reveal the general will, there must not be too much disagreement among citizens. "The more concord reigns in assemblies, that is to say the closer opinions come to unanimity, the more the general will predominates; whereas long debates, dissensions, disturbances, signal the ascendancy of particular interests and the decline of the State."[41]

The idea that some sort of democracy or collective choice rule might be the uniquely correct and determinate way for reasonable members of the public to resolve their moral disputes — by aggregating the individual judgments into a unique, rational shared judgment — has great appeal. If we all could will the law because it uniquely and rationally expresses our collective judgment, we could all will the same law, and so be both legislator and subject. But we have seen over the last fifty years that, as majority rule operates on great diversity of orderings, it tends to pathologies. This is not to say that, as a second-best or real-world device, majority rule may not be the best we can do;[42] rather, the fundamental point is that it is not an uncontroversial way to arrive at public moral norms given conditions of deep pluralism.

4.2 Rawls's Great (Early) Idea

Rawls was the first to see how the Kantian project of uncovering moral principles that can be legislated by all (and apply to all) can be cast in terms of a collective choice problem. As he notes in his seminal 1958 paper on "Justice as Fairness," we could try to derive the principles of justice "from *a priori* principles of reason, or claim that they were known by intuition."[43] Instead Rawls proposed to look at the choice of principles to govern social practices as a collective choice problem in which rational individuals *compromise* with each other when deciding on principles of justice.[44] Rawls was clearly aware how closely this project resembled certain problems in game theory. For now, I call attention to four points:

(i) A point of some interest (that is typically overlooked, especially by philosophers) is Rawls's remark that the reasoning of a party to the deliberative situation might be conceived of as "if he were designing a practice in which his enemy were to

---

[41] Rousseau, *The Social Contract*, p. 123 (Book IV, ch. 2)

[42] Through even as a device for actual political decisions I think majoritarianim is objectionable. See my "The (Severe) Limits of Deliberative Democracy as the Basis for Political Choice," *Theoria*, forthcoming.

[43] Rawls, "Justice as Fairness," p. 52.

[44] Ibid., p. 55

assign him his place."[45] It is not often noted that if this assumption were justified, maximin reasoning by the parties would be uncontroversially correct. This assumption would, essentially, make the parties' deliberations mimic reasoning in a zero-sum game, and, as Rawls well knew, von Neumann demonstrated that maximin is the correct solution to such games.[46] So *if* it was correct to see the choice problem in this way (which Rawls is driven to admit, it isn't), *then* the deliberative problem would have a determinate, uniquely rational, solution.

(ii) Rawls, however, did not pursue this justification of maximin. In "Justice as Fairness" he explicitly stated that the parts of game theory that most closely related to his project were cooperative games and group decision making, not zero-sum games.[47] It is remarkable that in 1958 Rawls already recognized that cooperative bargaining theory was relevant to his collective choice problem. Whereas Rousseau had proposed a collective aggregation solution to the problem of collective self-legislation for free and equal agents, Rawls began to develop a *bargaining solution*.

(iii) Rawls, however, rejected formal bargaining theory such as that proposed by R. B. Braithwaite in 1955. Rawls's objection — and this applies to other formal accounts such as John Nash's — is that threat advantage is relevant to the final bargain, and "To each according to his threat advantage is hardly a principle of fairness."[48] Thus, while Rawls clearly saw the choice problem as one that involved a sort of bargaining or compromise, he insisted that formal game theoretic approaches were inappropriate. The parties do not,

> as in the theory of games….decide on individual strategies adjusted to their respective circumstances in the game. What the parties do is to jointly acknowledge certain *principles* of appraisal relating to their common *practices* either as already established or merely proposed. They accede to standards to judgment, not to a given practice; they do not make any specific agreements, or bargains, or adopt a specific strategy. The subject of their acknowledgement is, therefore, very general indeed; it is simply the acknowledgement of certain principles of judgment, fulfilling certain general conditions to be used in criticizing the arrangement of common affairs….One could, if one likes, view the principles of justice as the "solution" of this highest order "game" of adopting, subject to the procedure described, principles of argument for all particular "games" whose peculiarities one can in no way foresee.[49]

---

[45] Ibid., p. 54. This remark, which has puzzled many commentators, is repeated in both editions of *A Theory of Justice*, p.133 (p. 152 of the 1971 edition).

[46] See *On Philosophy, Politics and Economics*, section 4.1.

[47] See "Justice as Fairness," note 9, which points the reader to these chapters of R. Duncan Luce and Howard Raiffa, *Games and Decisions* (New York: Wiley, 1957), viz. chs. 6, 14.

[48] Rawls, "Justice as Fairness," p. 58n.

[49] Ibid., p. 57.

(iv) Having rejected formal bargaining solutions, Rawls was left with two principles: equality and the Pareto-principle.[50] Equality, Rawls argued, would be accepted since "there is no way for anyone to win special advantage for himself."[51] (However, he also employed a version of maximin: since a practice that allows special treatment may turn against you, it is safer not to allow it.) The Pareto Principle was invoked as a defeater of the equality presumption: if some inequality-inducing improvement is preferred by everyone, then it will be agreed to. We thus get early formulations of the two principles: the first principle, which requires the greatest equal liberty, and the second, which allows inequalities that work to the advantage of all.

Formal bargaining solutions appear to give determinacy to the collective choice problem. Rawls, however, was quite right to reject them. Their determinacy is largely illusory: they yield clear determine solutions only if we accept their controversial frameworks. The most favored solution today is Nash's, but it can have counterintuitive implications. These led Braithwaite, in his application of game theory for moral philosophers, to advance an alternative bargaining solution. And, in *Morals By Agreement* David Gauthier relied on the Kali-Smorodinsky bargaining solution. Even disregarding this dispute, the determinacy is only at the level of mixes of cardinal utility satisfaction: until we specify the utility functions, the formal solution is of little help.

Without bargaining solutions the initial Rawlsian choice problem is indeterminate. The argument for egalitarian bargains is often a case of informal "splitting the difference" bargains, and, while often these bargains will arise, it is hard to see how, without a great many more assumptions, egalitarians bargains are the right general result.[52] The Pareto Principle is more solidly grounded as a principle of rational collective choice (if in everyone's ordering *Mx* is ranked as better than *My,* then *Mx* should be ranked as better than *My* in the social ordering). But, as Rawls came to realize, the Pareto Principle is highly indeterminate.[53] It seems that if one wishes to generate a collective moral deliberation situation with a determinate choice, one must specify the motivations and information sets of the parties in a much more detailed way. This, of course, is the path that led to *A Theory of Justice*: its strengths and weakness are well known. Instead of (to put the matter uncharitably) rigging the deliberative problem to give us a determinate result, let us explore the ignored other option: learning to live with the Pareto Principle's indeterminacy. *That is, let us consider*

---

[50] For an excellent analysis, see Robert Paul Wolff, *Understanding Rawls* (Princeton: Princeton University Press, 1977), chs. 4 and 5.

[51] Rawls, "Justice as Fairness," p. 55.

[52] See Binmore's complex argument for an egalitarian contract in *Natural Justice.*

[53] See Hardin, *Indeterminacy and Society*, ch. 4.

what our theory of a morality among free and equal persons will look like if we accept that the problem of collective legislation for members of P under C is inherently indeterminate.

5. Paretian Collective Deliberation

5.1 Unanimous Legislation I: The First Application of the Pareto Criterion

Recall our Kantian and Rousseauean-inspired problem: we seek moral requirements that each accepts as regulating their actions yet, since each gives it to herself, each remains free. Or, as Rawls puts it, we suppose that we have "free persons who have no authority over one another" and seek to determine whether there is a moral practice in which none are "forced to give into claims which they do not regard as legitimate."[54] However, in contrast to the early Rawls, we do not regard our deliberators as essentially self-interested. Our deliberators concur on a set of evaluative standards that all recognize as intelligible basis for deliberating about public morality, though they generally order these standards differently. Each consults her (intelligible to everyone) evaluative standards and, given them, proposes her version of the moral requirement to regulate some area of social life or some relevant practice. The set of all options is determined by the set of all proposed requirements. Given our assumptions of deep pluralism, different members of *P* under *C* will endorse differ rankings of moral requirements. This is precisely the point on which the Rousseauean-Rawlsian version of the problem differs from the ordinary interpretation of Kant: given rational evaluative diversity, we do not assume that the members of the public will propose the same requirements. In lieu of some powerful philosophical device such as an aggregation system, a bargaining theory, or a maximin motivation, there is no reason to think that all members of *P* under *C* converge on the same requirements, and every reason to think they will not.

Given the plural basis of the parties' deliberation, we cannot preclude that some members of *P* under *C* will propose requirements that others might find highly objectionable. To be sure, given that all are employing evaluative standards that all as members of *P* under *C* see as relevant to moral deliberation, there will not be out-and-out immoral or absurd proposals, such as "It required that all others be my slaves because that will be best for me," but our differences in evaluative standards can still lead some to endorse moral requirements that others find highly objectionable. Suppose we are deliberating about moral norms to regulate speech. Based on a ranking of not giving offense to others over freedom and other political values, a person may propose a highly restrictive doctrine according to which in all public speech, including political debate, one is morally prohibited from speaking in ways that any other citi-

[54] Rawls, "Justice as Fairness," p. 59.

zen considers offensive. To some free and equal moral person, such proposed moral requirement *Mz* may be worse than a full Hohfeldian liberty regarding political speech. If we all have such Hohfeldian moral liberties regarding speech each would have no moral duty refrain from any sort of speech, though no one would have a duty to refrain from interfering with the speech of others. At the point in a person's ordering at which she would place, on the basis of her evaluative standards, a Hohfeldian moral liberty over this area of social life to any remaining proposals, she has what we might call a "walk away point." She would rather walk away from collective deliberations than endorse such a requirement. There are, of course, generally great costs to this: a shared morality is in many ways fundamental to our social life and treating others as fellow moral persons. But we cannot insist that a person never has such a walk away point: there may be some sorts of moral requirements which she simply cannot see as legitimate.

Consider another case. Suppose that we are deliberating over schemes of property rights, and someone proposes a radical libertarian regime (with no proviso whatsoever that the regime must work out to everyone's advantage). Some might conclude that, given their evaluative standards, they would rank no system of property rights at all as superior. Of course there would still be some moral principles that would regulate property-related interactions (say, it would be wrong to invade people's rights to bodily integrity), but the objector may insist that no moral requirements establishing private property would be preferable to a system that, under the name of morality, excluded some from goods while making no commitment to universal benefit. So there is some point at which an individual would rather have no moral requirement regulating the practice than accept *Mz*. Again, suppose that within each individual's ordinal ranking of all the proposed principles, at some point the individual inserts "no moral requirement" at all. This bifurcates each individual's ordering into an eligible set (requirements that are better than Hohfeldian moral liberties) and an ineligible set (those that are ranked worse than moral liberties by a member of *P* under *C*).

5.2 Unanimous Legislation II: The Second Application of the Pareto Criterion

According to the first application of the Pareto criterion we can eliminate as a possible moral requirement among citizens of the realm of ends (members of *P* under *C*) any proposed moral requirement that is in the ineligible set of any member of the public. Only requirements that everyone holds are better than no requirements at all are in the eligible set. For us to appeal to a moral requirement outside the eligible set in our relations with the rejecter would, as Rawls says, to be insisting on standards of judgment that, as a free moral person, she cannot accept as legitimate: she cannot will them to be universal laws regulating all members of the public. One thing we might

mean by the inability to will a law — or its rational rejectability — is that no law at all would be better than such a law.[55]

We can invoke the Pareto criterion again: we can exclude any proposed requirement that, while in the eligible set of each individual, is Pareto-dominated by another proposed moral requirement. Requirement *My* is Pareto-dominated by *Mx* if and only if in each member of the public's ordering, $Mx \succ My$. If everyone holds that *Mx* is better than *My*, then the morality should be *Mx* rather than *My*. Acting on *My* would manifest a sort of collective irrationality: even though everyone sees it as inferior to *Mx*, we follow it anyway. This also might be seen as a sort of moral inefficiency, though moral philosophers may balk at the idea. What remains after our two invocations of the Pareto criteraion is a set of *optimal eligible* moral requirements: no proposed requirement in the set is ineligible in anyone's ranking, nor is it dominated by any other member of the set.

5.3 The Deliberative Model is Indeterminate

It has been the traditional aim of contractualist moral theory to whittle the set of optimal eligible requirements (over any area of social life or any practice) to a singleton. If we could design a choice situation among suitably described individuals such that one proposed requirement remained in the optimal eligible set, we would have discovered the uniquely correct moral duty. In this way moral philosophy could uncover the correct morality governing the realm of ends. As I indicated above, the move to a much thicker description of the choice situation in *A Theory of Justice* seemed motivated by such an aim. If we exclude "knowledge of those contingencies which set men apart…." then since "everyone is equally rational and similarly situated, each is convinced by the same arguments."[56] Thus the same set of requirements should be at the top of everyone's ranking. I believe that, along with the more noticed move to the political in Rawls's later work, he also abandoned the idea that only one set of principles of justice remained after the contractualist argument. Justice as fairness, as Rawls interpreted it in his later work, is simply one liberal conception of justice; because each of its constituent "elements can be seen in many different ways, so there are many liberalisms."[57] Rawls acknowledges that there are diverse interpretations of the basic concept of a liberal political order. Indeed, he insists that "it is inevitable and often desirable that citizens have different views as to the most appropriate political con-

---

[55] If we interpret the idea of a person being able to will law X as a member of *P* under *C* as implying that she does not think *any* other law is superior, then we will get a null set of universally-willed laws. This interpretation of being able to will a law is only plausible if we can justify a determinate deliberative solution — an idea I have argued we should abandon.

[56] Rawls, *A Theory of Justice*, pp. 17, 120.

[57] John Rawls, *Political Liberalism*, p. 223.

ception; for the public culture is bound to contain different fundamental ideas that can be developed in different ways."[58] Rawls also accepted that citizens arguing in good faith and employing public reason will not accept "the very same principles of justice."[59] Thus, in the end, Rawls tells us that the answer provided by public reason "must at least be reasonable, if not the most reasonable."[60] In his last work he abandoned the aspiration that the contractual argument reduces eligible conceptions of justice to a singleton.

As I think Rawls ultimately realized, the collective choice problem we have been discussing is indeterminate. Unless we invoke highly controversial notions such as a specific bargaining solution, or specify the evaluative criteria of the parties in great detail, we are left with a (non-empty) set of optimal eligible proposals.

6. Coordinating on a Morality

6.1 A 2x2 Toy Game Analysis

I am supposing, then, that the public justification of morality among members of the public leads, for every area of social life in which moral regulation is justified, to a set of optimal eligible interpretations that is not a singleton. Having taken rational collective self-legislation as far as we can go, we arrive at a number of possible sets of requirements (governing practices), all of which are evaluated as better than no moral regulation at all (i.e., pure Hohfeldian liberties), but none of which dominates the other. Now if morality was the sort of enterprise in which each could go her own way ("when it comes to being moral, I always say 'live and let live'") this would be no problem. But we have seen that to treat each other as free and equal moral persons we must concur on moral demands: we must accept common moral requirements. Each legislates for everyone. But, I have been arguing, you cannot have adequate reason to endorse a requirement that, in your own rational eyes, you have less reason to embrace than an alternative. Or can you?

At this point our members of the public face an impure coordination game along the lines of Display 1. Suppose that *X* and *Y* are alternative moral requirements regulating some practice. The numbers in the matrix refer to ordinal utility, with high numbers indicating highly ranked options; Alf's utility is in the lower left, Betty's in the upper right, of each cell. *It is crucial to stress that by "utility" here I simply mean a measure of the ranking of the options based on each person's evaluative standards. Utility here does not mean "self-interest" nor is it an independent value: it is simply a summary measure of how well an option satisfies the rational evaluative*

---

[58] Ibid., p. 227.

[59] Ibid., p. 214.

[60] Ibid., p. 246.

criteria of the individual qua member of P under C.[61] The uncoordinated outcomes indicate no moral practice at all on this issue (each is morally free to act as he or she wishes). Looked at e*x ante*, Betty's evaluative standards give her reason to endorse practice *X*; Alf's lead him to endorse *Y*. *Ex ante*, Betty does not have reason to endorse *Y* as legitimate over *X*, nor does Alf have endorse *X* rather than *Y*. They do, however, have reason to coordinate on either the *X* or the *Y* requirement.

		BETTY *X*	**BETTY** *Y*
ALF	*X*	3 2	1 1
	Y	1 1	2 3

Display 1: A Simple Impure Coordination Game
(3= best, 1 = worst)

Should Alf and Betty find themselves at *X,X* neither would have reason to change his or her action. Given each of their evaluative standards, they have the most reason to act on practice *X*. Should they instead find themselves at *Y,Y*, each will then have most reason (given his or her evaluative standards) to act on practice *Y*. Note that in neither case is the individual being induced by some external consideration to conform to a requirement that is not, from his or her perspective, optimal: c*onsulting simply his or her own evaluative standards, each has decisive reason to freely endorse whichever moral requirement they have coordinated on.* At morality *X* Betty can cite the requirements of *X* to Alf and, consulting only his own evaluative standards, he will have a reason to conform to *X*; and at morality *Y* Alf can cite *Y*, and Betty will have reason to act on it. And this even though, from the initial deliberative perspective, neither had reason to act on the other's preferred moral requirement.[62]

6.2 The Kantian Coordination Game: An N-person Iterated Toy Game

A one-shot two-person game can give us some insight, but it is clearly an inadequate way to model the selection of a moral requirement. The relevant coordination problem is not a single play game, but an iterated game. We have a number of encounters with others, and each can be understood as a play in a series of impure coordination games. Now in an iterated game a person's utility (again, remember this is defined

---

[61] This is fundamental point. I defend it in "Reasonable Utility Functions and Playing the Fair Way," *Critical Review of International and Social Philosophy*, forthcoming.

[62] Again, we should not be misled by the language of "preference." To prefer *X* to *Y* is simply to rank *X* over *Y* for purposes of choice; in our terms one's evaluative standards indicate reason to rank *X* over *Y* — this is all that is implied by saying one has a preference for *X* over *Y*.

solely in terms of her evaluative criteria) is a combination of her utility in this play, plus her expectations for utility in future games. Thus a person might sacrifice utility in one play to induce play in future moves that will yield her a more favored result. Moreover, it is certainly the case that in iterated games the play can move from one equilibrium to another. Peter Vanderschraaf and Brian Skyrms have shown how taking turns on each of the two equilibria emerges in iterated two-person games. [63]

However, in large *N*-person games with multiple coordination equilibria such solutions are much harder to sustain. In such large iterated games a bandwagon effect easily takes over. To intuitively see the driving force behind bandwagon effects, let us assume a cardinal utility measure (10 = best, 0 = no coordination) in a game with just two equilibria and nine players, as in Display 2:

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
X	2	3	4	5	6	7	8	9	10
Y	10	9	8	7	6	5	4	3	2

Display 2: Different Evaluations of Two Moral Requirements

If player A coordinates with another player on his preferred moral requirement (*Y*), he ranks that option as satisfying his evaluative standards to degree 10; if they coordinate on *X* he ranks the outcomes as 2. If he fails to coordinate — he acts on, say, *Y* while the other acts on *X*, they both get 0.

Now what is a member of *P* under *C* to do given these differences in evaluative standards? Consider a simple-minded but illustrative policy. Each begins play by endorsing, and acting upon, her favored option (except for player E who flips a coin and, given the flip, acts on the *Y* requirement). Again, if a player coordinates with another player on the same law, each gets their coordination payoff in Display 2; otherwise they receive 0 since they fail to coordinate. At the close of each round a player compares the score he received in that round with what he would have received if all others had played just as they did, but he played the opposite. If the opposite play would have resulted in a higher score, he changes his move. Assuming that each player meets every other player once in the first round, we have the following payoffs:

[63] Peter Vanderschraaf and Brian Skyrms, "Learning to Take Turns," *Erkenntis,* vol. 59 (2003): 311-46.

Partner→	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**Total**
Player A	–	10	10	10	10	0	0	0	0	40
Player B	9	–	9	9	9	0	0	0	0	36
Player C	8	8	–	8	8	0	0	0	0	32
Player D	7	7	7	–	7	0	0	0	0	28
Player E	6	6	6	6	–	0	0	0	0	24
Player F	0	0	0	0	0	–	7	7	7	21
Player G	0	0	0	0	0	8	–	8	8	24
Player H	0	0	0	0	0	9	9	–	9	27
Player I	0	0	0	0	0	10	10	10	–	30

Display 3: N-person Kantian Coordination Game, Round 1

In round 2, player F, given his own evaluative criteria, should switch his allegiance to *Y*; if F had played *Y* in round 1, he would have received 25 (5×5) rather than 21. Once F switches in round 2, at the end of round 2, G will find that he would have done better (24 rather than 16) by changing to *Y*, so G then will also change to *Y*. Obviously, once G also had changed to *Y*, H and I will also do so. We quickly reach an all-Y equilibrium.

6.3 The Increasing Returns of Shared Moral Requirements

The Kantian Coordination Game is, of course, still terribly over-simplified, depending on a rather dumb decision rule, and an assumption that all players meet all others an equal number of times.[64] And of course we have supposed a certain population distribution. It is by no means inevitable that the public must converge on a common convention. If in Display 2 the entire population was evenly divided between A-type and I-type utility functions, the population could easily settle into a "polymorphic" equilibrium, with A-types always playing *Y* and I-types always playing *X*. Note that this is more likely to occur with populations split entirely into radically opposing groups and where each group ranks the other's alternative as only marginally better than no coordination at all.[65]

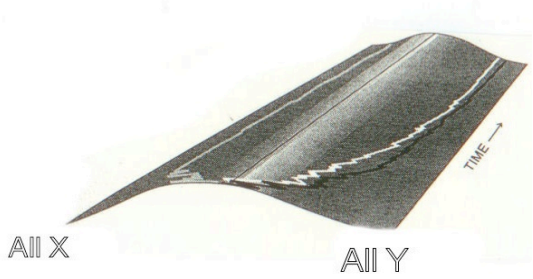
---

[64] As Brian Skyrms shows, if players can detect other players with complementary utility functions, the analysis of the game is very different. See his *Evolution of the Social Contract* (Cambridge: Cambridge University Press, 1996), ch. 1. There has not been a great deal of work modeling what equilibrium will emerge in iterated impure coordination games; some experiments cast doubt on whether any simple mechanism, such as the most "salient" solution, will be adopted. See Morton D. Davis, *Game Theory* (Mineola, NY: Dover, 1983), pp. 133-35. On uncertainty in coordination games, see Fernando Vega-Rodondo *Economics and the Theory of Games* (Cambridge: Cambridge University Press, 2003), pp. 188ff.

[65] This raises the interesting possibility of a Kantian account of moral multi-culturalism.

Despite its obvious limitations, The Kantian Coordination Game brings out a crucial feature of moral life among free and equal persons with a commitment to respecting each other's status: the increasing returns of coordinating on a common understanding of moral requirements. We can think of each member of *P* under *C* as having two distinct morality-related desiderata: (1) to act on the moral requirement that best satisfies her evaluative standards and (2) to act on moral requirements that are embraced by all, so that in her interactions she can make moral demands that respect the equality and moral freedom of all. Other things equal, a member of *P* under *C* has reason to seek a common moral life that conforms to (1), but as more and more other free and equal persons come to act on some member of the optimal eligible set, the second desideratum comes increasingly into play. Coming to accept the moral requirements that others do, so long as it is in the optimal eligible set, comes to be the actual way in which each member of the public can best satisfy her entire set of evaluative standards.

Formally, converging on a common morality is an instance of increasing returns: the more others come to embrace a certain moral requirement, the more reasons others have to also embrace it.[66] As we see in Display 2, some people's evaluative standards may strongly favor an alternative moral requirement (consider persons A and I), yet so long as everyone places significant importance on acting as others do (the second desideratum), our members of the public can still end up coordinating: as more and more adopt an alternative, even those who strongly favor another option come on board. As one option (perhaps simply because of some random event) becomes slightly more popular than the others, people will gravitate to that option (as it stands the best chance of universal acceptance), and we witness a "bandwagon" effect based on the increasing returns for everyone of adopting the more poplar option. This dynamic is illustrated in Display 4.



Display 4: Increasing Returns Dynamics (drawn from Arthur, *Increasing Returns and Path Dependency in Economics*, p. 3)

[66] The path-breaking work on increasing returns was done by W. Brian Arthur. See his *Increasing Returns and Path Dependency in Economics* (Ann Arbor: University of Michigan Press, 1994).

As we can see, starting out with a population evenly split between advocates of *X* and of *Y*, random events can lead the population to all *X* or all *Y* equilibria. Which equilibrium emerges will be path-dependent: at time zero there is no reason why one or the other should emerge as the *unanimously-selected choice*. Chance events, people's reactions to what they perceive as the favored option, the publication of *A Theory of Justice* in 1971 — any can lead an idealized population of Kantians to converge on one member of the eligible set. But once we have arrived at such a convergence, each member of the public, consulting only her own evaluative standards, will freely act on the chosen moral requirement. For our purposes what is crucial is that the contingent and accidental way in which large groups can come to coordinate on a common practice is no bar to there being a determinate morality that all can endorse given their evaluative criteria *once it has been arrived at*.

7. The Implications of the Analysis

That our Kantians could come to share common moral requirements through iterated coordination games — or more generally convergence over time because of increasing returns dynamics — does not, of course, show us that our social morality actually evolved in this way. Insofar as having a common morality is necessary to treat others as equal moral persons in one's daily interactions, the dynamics I have been considering are part of an adequate account of how we have come to share a morality, but it would be pressing credulity to think that this is the complete story. Some may go further and insist that it isn't even an important part of the story: why we have actually come to have certain moral practices and rules, they will say, depends on biological evolution, social power and a host of other hard-headed concerns, not something so ideal as respecting others. This sort of hard-headedness seems more appealing at first sight than after reflection: that we are concerned with how our moral claims appear to others, and whether they can see a reason to abide by them, is probably a far more important factor in moral thinking than we are first apt to think. Unless the requirements of morality are affirmed by the reason of most people, it is unlikely in the extreme that a society's moral order will be stable over the long-run.

The main implications of the analysis, however, do not concern the explanation of how we have arrived at our morality, but our understanding of what moral theory is, and what is demanded by requirement that we respect others under conditions of deep evaluative plurality. Today, I think, we tend to think of moral theory and rational reflection as seeking to provide determinate answers to what morality requires. We first reflect on what a rational justified morality is and then examine our actual morality to see if it measures up. The history of thinking in this way gives us ample cause to

doubt whether such rational determinacy is to be had. We have witnessed in the last thirty or so years a plethora of normative theories, each giving determinate but widely diverging pronouncements about the content of our *bona fide* moral requirements. I have suggested that there is good reason to conclude that, under conditions of deep evaluative pluralism, the idea of impartial rational reflection is indeterminate. Rational reflection can narrow the field, but actual interactions of good-willed people are needed to fill in the large gaps, and give us a morality that we all can will.

Once we realize that arriving at a fully justified morality could — indeed must — involve chance and path-dependency, we are apt to see moral theory in a different light. In the view of a previous generation of moral philosophers such as P.F. Strawson and Kurt Baier,[67] the starting place of moral philosophy is our actual moral practices. The question for the moral philosopher is: can these actual moral practices be justified as ones that would be acceptable from the impartial moral point of view? In our terms, the task of the moral philosopher is to determine whether our current moral practices are in the optimal eligible set: that is the best (and it is quite a bit) that impartial rational reflection can do.

To respect others as free and equal persons does not require that we show the moral demands that we make on them are uniquely rational, or are, from their perspectives, the moral demands that best conform to their evaluative standards. Because so many moral philosophers have thought that respect must require this, they have either sought to ignore the extent of evaluative pluralism (if we all value the same thing, our rational judgments *must* converge) or invent powerful philosophical devices that (miraculously?) take our diverse evaluative judgments as inputs and yield a single, uniquely rational, determinate, answer. As philosophers we enjoy such constructions (and finding inevitable flaws), but the supposition that respecting others as free and equal requires such unequivocal answers generated by controversial devices is ultimately morally corrosive. The plausible lesson many draw from these repeated failed attempts is that respecting all as free and equal must ultimately be impossible. A moral theory that justifies our current practices if they are eligible moral requirements has a more modest ambition, but fulfilling it is all that is needed to dispel the fear that our moral demands might be just a way of pushing others around.

Philosophy
University of Arizona

---

[67] I have in mind here P. F. Strawson, "Social Morality and Individual Ideal," *Philosophy*, vol. 36 (Jan., 1961): pp. 1-17; Kurt Baier, *The Moral Point of View: A Rational Basis for Ethics*, abridged edn. (New York: Random House, 1965).