

The Demands of Impartiality and the Evolution of Morality*

Gerald Gaus

I. 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 Rawls under the proper conditions collective deliberation by rational and reasonable parties could yield agreement on impartial rules of justice.

From another point of view moralities are social facts with histories. The heroes of this tradition are Hume, Ferguson and, Smith. 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.

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. Section 4 briefly considers Rawls's early proposal for solving the problem of public justification under evalua-

* I have presented my thoughts on these matters to a number of audiences, and have greatly benefited from their questions and comments. I would like to 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, the Workshop on Equal Respect for Persons held at the University of Genova, the Philosophy Departments at the University of Reading, North Carolina State University, the University of Kansas, and 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.

tive pluralism. Section 5 points the way to a more adequate approach to the problem, but we shall see that the solution is indeterminate; Section 6 argues that social evolutionary processes can complete the justificatory process. I reflect on some of the implications of the analysis in Section 7.

2. RESPECT FOR PERSONS AND THE IMPARTIAL AUTHORITY OF MORALITY

2.1 *Morality, Authority and the Threat of Subjugation*

Social morality provides a set of principles that allows a 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:² 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.”³ To make a moral demand on another is to assume a practical authority over another to make demands and to demand compliance.⁴ 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.⁵ 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 so what she should do. If she does not comply, 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.”⁶ 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 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

² 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.

³ Stephen Darwall, *The Second-person Standpoint: Morality, Respect and Accountability* (Cambridge, MA: Harvard University Press, 2006), p. 4.

⁴ *Ibid.*, pp. 10–11.

⁵ *Ibid.*, p. 76.

⁶ Thomas Hobbes, *Leviathan*, Michael Oakeshott, ed. (Oxford: Basil Blackwell, 1948), p. 180 (ch. 26).

that in their own controversies: bewraying [sic] their want of right reason by the claim they lay to it.⁷

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 stability. Because of *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."⁸

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."⁹

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.¹⁰

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.¹¹ 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.¹² 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,¹³ Mill's own accep-

⁷ Ibid., p. 26 (ch. 5).

⁸ Jeffrey Reiman, *Justice and Modern Moral Philosophy* (New Haven, CT: Yale University Press, 1990), p. 1.

⁹ Robert Filmer, *Patriarcha* in Peter Laslett, ed., *Patriarcha and Other Political Works* (Oxford: Blackwell, 1949), p. 53.

¹⁰ Filmer, "Directions for Obedience to Government in Dangerous or Doubtful Times," in *ibid.*, p. 231.

¹¹ Filmer, *Patriarcha*, pp. 73-74.

¹² Ibid. p. 96.

¹³ Aristotle, *Nicomachean Ethics*, Sir David Ross, trans. (Oxford: Oxford University Press, 1954), p. 212 [1161a30-b19].

tance of authoritarianism for “races” in their “nonage,”¹⁴ 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*¹⁵ — 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.¹⁶

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. The authority of morality cannot be partial: it cannot privilege the perspective of some free and equal more person over another’s such that one simply occupies the role of legislator while the other is subject. How can liberalism’s commitments to moral freedom and the absence of natural authority of one person over another 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.¹⁷

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.

Kant’s depiction of the self-legislative nature of a free and impartial 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,¹⁸ 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*.”¹⁹ 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.

2.3 *The Generic Public Justification Principle*

If we take seriously the unanimity requirement implicit in Kant’s notion of universal legislation, we are led to a view of an impartial moral justification along the lines of:

¹⁴ Mill, *On Liberty*, ch. 1, para. 10.

¹⁵ Henry Sidgwick, *The Methods of Ethics*, 7th edn. (Chicago: University of Chicago Press, 1962), pp. 489ff.

¹⁶ For a general characterization of moral authoritarianism, see my *Social Philosophy* (Armonk, NY: M.E. Sharpe, 1999), pp. 6ff.

¹⁷ Immanuel Kant, *Foundations of the Metaphysics of Morals*, Lewis White Beck, ed., trans. (Indianapolis: Bobbs-Merrill, 1959), p. 52. [Akademie 434]

¹⁸ Ibid, pp. 51-52 [Akademie, 433-34].

¹⁹ Ibid.

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.

The Public Justification Principle, as Rawls puts it, conceives of impartial moral principles as mutually acknowledged “by free persons who have no authority over one another.”²⁰

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 .²¹ 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; Rawls’s parties are reasonable and rational.

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 requirements 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.

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.²² Suppose we understand a member i of P under C as consulting her relevant evaluative standards — the full set of considerations that is relevant to her decision whether to accept some moral requirement (§3). After consulting her evaluative standards, i proposes her preferred moral requirement, M_i : the moral requirement that, on her (somewhat idealized) reasoning, best conforms to her evaluative standards. (This procedure is akin to that

²⁰John Rawls, “Justice as Fairness” in Samuel Freeman, ed., *John Rawls: Collected Papers* (Cambridge, MA: Harvard University Press, 1999), p. 55.

²¹It 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 the inadequacy of an earlier formulation.

²²John Rawls, *A Theory of Justice*, revised edn. (Cambridge, MA: Belknap Press of Harvard University Press, 1999), p. 16 (p.17 of the original edition).

utilized by Rawls in “Justice as Fairness.”)²³ Suppose also that, on the basis of her 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 of making the choice problem determinate must lead to 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 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.”²⁴ 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

²³ “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.

²⁴ Rawls adds: “within the framework of free institutions of a constitutional regime.” *Political Liberalism*, paperback edn. (New York: Columbia University Press, 1996), p. xviii.

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_1 over M_2 .²⁵ 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 members of P draw on to devise proposals and ranking of moral requirements.

We suppose, then, plurality of evaluative standards for P under C . *But how great is this pluralism?* Again, each public reason view will specify a different characterization. Any plausible liberal public reason view, however, must admit (i) great diversity of evaluative standards (and so recognize the importance of reasonable pluralism) while (ii) also limiting the range of considerations that may be drawn upon in justification. These limits (point b), I think, are implicit in the very idea of public justification. The point of public justification is for Alf to treat Betty as a free and equal moral person while also demanding that she conform to certain moral requirements. He can do this if, from her own evaluative perspective, she too has reason to accept these requirements. Suppose, then, Alf seeks to justify requirement M to Betty by appealing to her standard Σ_B , which leads her to endorse M . But suppose that Alf also holds that Σ_B is not an intelligible or reasonable basis for endorsing M ; on his view, either she has no good reason to hold Σ_B , or there is no sound deliberative route from it to M . If so, then he cannot understand himself to have justified M . As a member of the public, he cannot think that deliberation based on that standard provides Betty (as another member of the public) with a reason to endorse a requirement if in his view it is an unintelligible or unreasonable basis for her deliberation. That her unreasonable standard leads her to accept M cannot lead him to think Betty has a reason to endorse M : garbage in, garbage out. A plausible conception of evaluative pluralism then, must accept some version of what we might call “mutually intelligible evaluative pluralism” *at the level of members of P*. Members of P will see themselves as deeply disagreeing about the basis for accepting a requirement, but will acknowledge that the bases of others’ reasoning is intelligible and is relevant to the justificatory problem. As Isaiah Berlin might say, the range of plausible pluralism of members of P is limited by the “common human horizon.”²⁶ Moreover, there is empirical evidence that our actual value disagreements are not so much about what is or is not valuable, but how we order shared (and so mutually intelligible) values.²⁷

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 re-

²⁵ 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.

²⁶ See my *Contemporary Theories of Liberalism: Public Reason as a Post-Enlightenment Project* (London: Sage, 2003), ch. 2.

²⁷ 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.

sult in great disagreement in their rankings of candidates for moral requirements. If a member of the public Alf holds ranking $\Sigma_1 > \Sigma_2$ (read as “ Σ_1 is ranked above Σ_2 ”) while Betty maintains that $\Sigma_2 > \Sigma_1$, then if these are the only relevant standards, and, if within a perspective the degree of justification of moral requirements varies with the ranking of evaluative standards, Alf will hold $M_1 > M_2$, while Betty 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 members’ of P 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 rational consensus, but evaluative pluralism leads to disagreement. What’s a liberal to do?²⁸

4. RAWLS’S GREAT IDEA

Because we have attributed significant evaluative pluralism to our members of the public in the deliberative model, the Kantian project of uncovering moral principles that can be legislated by all (and apply to all) is best cast in terms of a collective choice problem. Rawls was the first to see this. 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.”²⁹ 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.³⁰ 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 in the deliberative situation might be conceived of as “if he were designing a practice in which his enemy were to assign him his place.”³¹ It is seldom appreciated that if this assumption was 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.³² So *if* it was correct to see

²⁸ 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.

²⁹ Rawls, “Justice as Fairness,” p. 52.

³⁰ *Ibid.*, p. 55.

³¹ *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).

³² See *On Philosophy, Politics and Economics* (Belmont, CA: Wadsworth, 2007), section 4.1.

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.³³ It is remarkable that in 1958 Rawls already recognized that cooperative bargaining theory was relevant to his collective choice problem. Rawls thus 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.”³⁴ 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.³⁵

Formal bargaining solutions appear to give determinacy to the collective choice problem. Their determinacy, though, 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.

(iv) Having rejected formal bargaining solutions, Rawls was left with two principles of choice: equality and the Pareto-principle.³⁶ Equality, Rawls argued, would be accepted since “there is no way for anyone to win special advantage for himself.”³⁷ (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

³³ 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.

³⁴ Rawls, “Justice as Fairness,” p. 58n.

³⁵ *Ibid.*, p. 57.

³⁶ For an excellent analysis, see Robert Paul Wolff, *Understanding Rawls* (Princeton: Princeton University Press, 1977), chs. 4 and 5.

³⁷ Rawls, “Justice as Fairness,” p. 55.

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.

Because Rawls rejected formal bargaining solutions his choice problem was indeterminate. The argument for egalitarian bargains is often a case of informal “splitting the difference” bargains, and, while these bargains will sometimes arise, it is hard to see how, without a great many more assumptions, egalitarian bargains are the right general result.³⁸ The Pareto Principle, however, is much more solidly grounded as a principle of rational collective choice (if in everyone’s ordering M_1 is ranked as better than M_2 , then M_1 should be ranked as better than M_2 in the social ordering). But, as Rawls came to realize, the Pareto Principle is often indeterminate.³⁹ If we wish to generate a collective moral deliberation situation with a determinate choice, we must specify the motivations and information sets of the parties in a detailed way so that they all reason the same and will choose the same point on the Pareto frontier. Thus the path that led to *A Theory of Justice*: its strengths and weakness are well known. Although Rawls began by posing a problem of collective choice, ultimately, as he tells us, the problem is reduced to the reasoning of a single person. 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.”⁴⁰ Instead of (to put the matter uncharitably) rigging the deliberative problem to give us a determinate result, let us explore the ignored option: learning to live with the Pareto Principle’s indeterminacy. *That is, let us consider 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

Let us now return to the companion deliberative model. 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 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 citizen considers offensive. To some free and equal moral person, such a proposed moral requirement M_x may be worse than a full Hohfeldian liberty regarding political speech. If we all have Hohfeldian moral liberties re-

³⁸ This is brought out by Ken Binmore’s complex argument for an egalitarian contract in *Natural Justice* (Oxford: Oxford University Press, 2005),

³⁹ Russell Hardin, *Indeterminacy and Society* (Princeton: Princeton University Press, 2003), ch. 4.

⁴⁰ Rawls, *A Theory of Justice*, pp. 17, 120.

garding 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. Now at the point in a person's ordering at which she would place, on the basis of her evaluative standards, a Hohfeldian moral liberty — a “no moral requirement” — over this area of social life to any remaining proposals, she has what we might call a “no agreement point.” She would rather not have a collective agreement 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 must hold that every proposal is better than failure to legislate: there may be some proposed moral requirements that she simply cannot see as in any way legitimate. We thus bifurcate 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, 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.⁴¹

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 M_2 is Pareto-dominated by M_1 if and only if in each member of the public's ordering, $M_1 > M_2$. If everyone holds that M_1 is better than M_2 , then the morality should be M_1 rather than M_2 . Acting on M_2 would manifest a sort of collective irrationality: even though everyone sees it as inferior to M_1 , we follow it anyway. What remains after our two invocations of the Pareto criteria 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 cor-

⁴¹ If we interpret the idea of a person having reason to accept M as a member of P under C as implying that she does not think that there is *any* superior alternative requirement, then we will get a null set of “universally-willed requirements.” This interpretation of universal legislation is only plausible if we can justify a determinate deliberative solution — an idea I have argued we should abandon.

rect 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 the aim of ensuring that the same requirement 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."⁴² 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 conception; for the public culture is bound to contain different fundamental ideas that can be developed in different ways."⁴³ Rawls also accepted that citizens arguing in good faith and employing public reason will not accept "the very same principles of justice."⁴⁴ Thus, in the end, Rawls tells us that the answer provided by public reason "must at least be reasonable, if not the most reasonable."⁴⁵ 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. 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, 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.

At this point our members of the public face an impure coordination game along the lines of Display 1. Suppose that M_1 and M_2 are alternative moral requirements in the optimal eligible set. 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 evaluative criteria of the indi-*

⁴² John Rawls, *Political Liberalism*, p. 223.

⁴³ *Ibid.*, p. 227.

⁴⁴ *Ibid.*, p. 214.

⁴⁵ *Ibid.*, p. 246.

*vidual qua member of P under C.*⁴⁶ The uncoordinated outcomes indicate no moral requirement at all on this issue (each is morally free to act as he or she wishes). Looked at *ex ante*, Betty's evaluative standards give her reason to accept practice M_1 ; Alf's lead him to accept M_2 . *Ex ante*, Betty does not have reason to accept M_2 over M_1 , nor does Alf have accept M_1 rather than M_2 . They do, however, have reason to coordinate on either of the two requirements rather than none at all.

		BETTY	
		M_1	M_2
ALF	M_1	3	1
	M_2	1	2

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

Should Alf and Betty find themselves at M_1/M_1 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 M_1 . Should they instead find themselves at M_2/M_2 , each will then have most reason (given his or her evaluative standards) to act on M_2 . Note that in neither case is any party induced by some external consideration to conform to a requirement that is not, from his or her perspective, optimal: *consulting simply his or her own evaluative standards, each has decisive reason to freely endorse whichever moral requirement they have coordinated on*. At M_1/M_1 Betty can demand that Alf conform and, consulting only his own evaluative standards, he will have a reason to conform; and at M_2/M_2 Alf can demand M_2 , 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.⁴⁷

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 from the optimal eligible set. 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 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

⁴⁶ This is fundamental point. I defend it in "Reasonable Utility Functions and Playing the Fair Way," *Critical Review of International and Social Philosophy*, 2009.

⁴⁷ Again, we should not be misled by the language of "preference." To prefer M_1 to M_2 is simply to rank M_1 over M_2 for purposes of choice; in our terms one's evaluative standards indicate reason to rank M_1 over M_2 — this is all that is implied by saying one has a preference for M_1 over M_2 .

Vanderschraaf and Brian Skyrms have shown how taking turns on each of the two equilibria emerges in iterated two-person impure coordination games.⁴⁸

However, in large N -person impure coordination games with multiple equilibria such solutions are, I think, practically impossible. In such large iterated games a bandwagon effect 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
M_1	2	3	4	5	6	7	8	9	10
M_2	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 (M_2), he ranks that option as satisfying his evaluative standards to degree 10; if they coordinate on M_1 he ranks the outcomes as 2. If he fails to coordinate — he acts on, say, M_2 while the other acts on M_1 , they each 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 employing her favored requirement (i.e., making demands based on it) in all her interactions (except for player E who flips a coin and, given the flip, acts on the M_2 requirement). Again, if a player coordinates with another player on the same requirement, each gets her coordination payoff in Display 2; otherwise each player receives 0 since they fail to coordinate. At the close of each round a player compares the score she received in that round with what she would have received if all others had played just as they did, but she played the opposite. If the opposite play would have resulted in a higher score, she changes her move. Assuming that each player meets every other player once in the first round, we have the following payoffs:

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 M_2 ; if F had played M_2 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 she would have done better (24 rather than 16) by changing to M_2 , so G then will also change to M_2 . Obviously, once G also had changed to M_2 , H and I will also do so. We quickly reach an all- M_2 equilibrium.

⁴⁸ Peter Vanderschraaf and Brian Skyrms, “Learning to Take Turns,” *Erkenntnis*, vol. 59 (2003): 311-46.

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.⁴⁹ 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 settle into a “polymorphic” equilibrium, with A-types always playing M_2 and I-types always playing M_1 . 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.⁵⁰

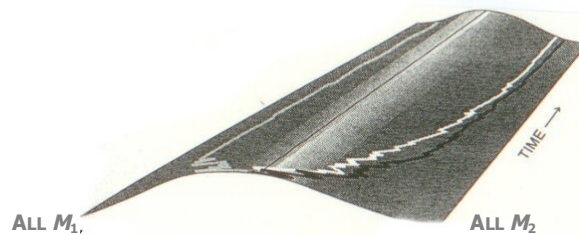
Despite its obvious limitations, The Kantian Coordination Game brings out a crucial feature of moral life among free and equal persons with a commitment to respecting each other’s status: the increasing returns of coordinating on a common understanding of moral requirements. We can think of each member 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 reason others have to also embrace it.⁵¹ As we see in Display 2, some people’s evaluative standards may strongly favor an alternative moral requirement (consider person 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 popular option. This dynamic is illustrated in Display 4.

⁴⁹ 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.

⁵⁰ This raises the interesting possibility of a Kantian account of moral multi-culturalism.

⁵¹ 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).



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

As we can see, starting out with a population evenly split between advocates of M_1 and of M_2 , random events can lead the population to all M_1 or all M_2 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 implication of the analysis, however, does 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 the 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 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,⁵² 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 from the impartial perspective. Because so many moral philosophers have thought that respect requires 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

⁵² 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).