

# Moral Conflict and Prudential Agreement: Michael Moehler's *Minimal Morality*

*Gerald Gaus*

Michael Moehler's *Minimal Morality* is a wonderful and important book, from which I have learned a great deal. It reinvigorates rational choice moral theory in the process of confronting what I see as the most important issue in social and moral philosophy today: can those in a deeply morally divided society endorse a common moral framework to structure social cooperation? Is a rational moral order possible under conditions of deep and wide moral diversity? *Minimal Morality's* answers are thoughtful and innovative. I am not, however, entirely persuaded that *Minimal Morality* can fulfill its promise of grounding a cooperative order when "traditional" modes moral justification cannot.

## *1. Two "Traditional" Approaches to Moral Justification Under Diversity*

*Minimal Morality* is offered as a framework for cooperation when social conflicts are so deep that public reason approaches to moral justification fail. Consider two reasonably well-known public reason approaches. As we know, Rawlsian public reason is offered as a response to the fact that under free institutions the exercise of human reason leads to "intractable struggles" and "irreconcilable" conflicts of "absolute depth" about the nature of the universe, the world, value and justice (Rawls, 2005: xxvi, 4). To cope with such diversity Rawls and his followers seek to identify some basic, shared, moral commitments — some minimal level of agreement — that will secure a shared moral framework for cooperation. Let us call this the *Shared Moral Basis Approach*. On this approach the search for a public morality in a deeply divided world cannot be based, as it were, on *moral diversity all the way down*: diversity must bottom out in basic agreement on some shared values on which a shared moral framework can rest.

The obvious worry about the Shared Moral Basis Approach is that its prospects for success are inversely related to the extent of moral diversity: the wider and deeper the moral diversity in our society, the less prospect there is for finding a suitable shared basis. It is for this reason that Rawls (2005: xlvii), while admitting that the underlying factors that lead to reasonable pluralism about the good also characterize reasoning about justice, nevertheless

insisted that they are less serious in the case of justice. The extent of diversity about justice is less — and far more manageable — than diversity about the good. Rawls believed this shared basis was to be found in the basic values of liberal democratic political culture.

Once he recognized reasonable disputes about justice — and thus that his preferred doctrine of “justice as fairness” could be reasonably rejected — Rawls’s (2005: xxxvi, xlvi) proposals become much vaguer, focusing on a broad “family” of liberal doctrines. But dispute about the content of shared values, and so the nature of the “family,” has emerged (e.g., Tomasi, 2012). Partly in response to increasing doubts as to whether there is a sufficiently thick moral consensus for Rawls’s project to succeed, I (Gaus, 2011: 283-92) and others (Vallier, 2014: 103-40; Muldoon, 2016) have proposed what might be called the *Diverse Moral Inputs Approach*, the so-called convergence theory of public reason. Here, the aim is to see whether, even if diversity of moral reasoning goes all the way to different foundations and commitments, diverse agents can nevertheless endorse the same rules of social morality. That we start in different places does not mean that we cannot end up on common ground. This approach does not identify a canonical set of reasons that all share, but seeks to understand social dynamics that can lead those who disagree all the way down on moral foundations to nevertheless come to a common endorsement of the moral framework. As Brian Kogelmann (2017) has nicely put it, rather than going far *down* to find agreement, we look *up* — to the moral rules that diverse agents can agree upon.

Critics of the Rawlsian Shared Moral Basis Approach have a relatively easy task: they simply query whether the values identified by Rawls — say, the two moral powers — are as widely shared as Rawls claims, or whether they lead to shared rules of justice. So for Moehler to show that his theory helps out when disagreement is too deep for the Shared Basis Approach, he need simply show that some societies do not share the requisite shared beliefs. On the other hand, showing that moral conflict is too deep for the Diverse Moral Inputs Approach is rather more involved. Rather than simply showing that the minimal basis beliefs ( $\beta_1 \dots \beta_n$ ) are not shared in some society *S*, Moehler must argue that in *S* the forces that lead diverse moral agents to converge on common rules are weaker than the forces that lead them into conflict. Because the Diverse Moral Inputs Approach does not identify the set of agreed-upon reasons for direct evaluation of how widely their content is shared, but instead offers an account of the dynamics of convergence, a critic must hold that the gravitational pull towards endorsing the same basic moral rules is overwhelmed by centrifugal forces of

diversity. Our moral disagreements, the critic must hold, are stronger than our moral concern with sharing common rules of justice — we rank moral conflict over sharing moral rules and the cooperative order they support.

## 2. *An Ultraminimal Shared Basis View*

Moehler holds that his Minimal Morality could still obtain under these conditions of deep and wide moral conflict. Indeed the very point of Minimal Morality is to provide what we might call a “back-up” morality: a route to moral order when moral diversity undermines the ability of the devices of so-called “traditional morality” to justify moral rules. Focusing on the relation of his analysis to the Diverse Moral Inputs Approach, he is committed to:

*The Hobbesian Claim:* When the centrifugal forces of moral conflict are too strong to sustain shared rules of justice or social morality, Minimal Morality can provide a basis of minimal moral order by securing peace.

Moehler’s defense of the Hobbesian claim is what I shall call an *Ultraminimal Shared Basis Approach*. It is, as I see it, an ultra-weak version of a Shared Basis view insofar as, at the end of the day, the account *does* find a robust shared value from which we can derive the conditions for order: if we go far enough down we find that *prudence leads us to the terms of a stable cooperation*. Even in cases where the centrifugal forces of moral conflict are too strong to sustain common rules of justice, prudent individuals share deep reasons on which they can erect a minimalist, peaceful, moral order.

I begin, then, with this Ultraminimal Shared Basis: prudence. Moehler writes:

*Homo prudens* is a forward-looking agent who, despite that she may discount future benefits according to her time preference, values her life together with the expected gains from peaceful long-term cooperation more than noncooperation per se in any particular case of conflict (17).<sup>1</sup>

*Homo prudens* greatly values her life, autonomy and the satisfaction of her basic needs (116) but more than that, conjoined with “the expected gains from peaceful long-term cooperation,” *her values always favor pursuing peace over serious conflict*. Moehler is careful to identify a specific type of conflict — conflict in the “strict” sense — which his theory is meant to address: It is characterized by two features (15) (the names are mine):

---

<sup>1</sup> Unless otherwise noted, parenthetical page references refer to Moehler (2018).

1. *The Centrifugal Condition*: “the parties to the conflict have tried to resolve their disputes based on the particular moral ideals that they embrace and their implicitly or explicitly agreed moral conventions that may be historically justified or the product of biological, social, or cultural evolutionary processes.”

This condition identifies the conflict as one that cannot be resolved by “traditional” morality (such the Diverse Moral Inputs Approach) and so we need a back-up morality: the centrifugal forces of moral disagreement overwhelm the tendency to converge on shared moral rules. Note, though, that the Centrifugal Condition is open to a weak and a strong interpretation:

1a. *The Strong Centrifugal Condition*: the parties to the conflict have tried to resolve their disputes based on their particular moral ideals ... and their continued moral conflict puts them deeply at odds.

1b. *The Weak Centrifugal Condition*: the parties to the conflict have tried to resolve their disputes based on their particular moral ideals ... and while their moral views provide no help in resolving their conflict, neither do they greatly exacerbate it.

I think the difference between these is important, a point to which I shall return. The second feature specifying the nature of the conflict is:

2. *The Conflict Condition*: “the parties are so severely negatively affected by the points of contention that they are prepared to engage in destructive actions, including actions that endanger the lives of their conflict partners or those who are close to them, if the disputes remain unresolved, because the agents consider destructive actions to be more beneficial to them than remaining in their current situations” (15).

People are prepared to fight rather than submit to the status quo. This condition too is not quite as straightforward as it may at first appear. The Conflict Condition identifies a general set of “destructive” actions which includes, but is by no means restricted to, acts that endanger lives. Now just what actions are “destructive” (as opposed to honorable or exhilarating) is to some extent internal to a moral view. Hobbes, I think, held that glory seekers may positively value conflict — and this was a general theme of Romantics, such as the “Storm and Stress” movement of eighteenth century Germany. Moreover, individuals will certainly disagree about the specific contexts in which peaceful resolution is better than conflict. *Pace* the social contract tale, we never simply choose global order or the state of nature, but cooperation here and conflict there. Even during intense civil conflict cooperation on some matters is often maintained. In the American Civil War, from 1862-63, Union and

Confederate armies “paroled” prisoners of war on the condition that they promised not to resume fighting once released (Macpherson, 1988: 791). We should resist any model that would confront us with a binary choice of all war or all cooperation.

In any event, the core of the *homo prudens* model is that when the Centrifugal and Conflict conditions are met, purely instrumental Minimal Morality is a rational response. According to Moehler:

The only basis for the justification of a principle of conflict resolution for cases of conflict in the strict sense defined is that agents are rational, *they share a similar human nature and empirical environment, and they have an overarching interest in securing peaceful long-term cooperation* (17, emphasis added).

Here we have a clear indication that this is an Ultraminimal Shared Basis view: we share this rock-bottom interest in peaceful cooperation. Underlying all our moral conflict is, after all, a basis for Minimal Morality in a shared human nature — *homo prudens*. Recall again the critical claim: “*Homo prudens* ... values her life together with the expected gains from peaceful long-term cooperation more than noncooperation per se in any particular case of conflict” (17). What *homo prudens* would agree to in a suitably described choice situation essentially specifies Minimal Morality.

### 3. *When Morality Drives Us Apart*

Now I think this Hobbesian strategy is plausible for cases described by The Weak Centrifugal Condition. Leaving aside disagreements about when a rule of cooperation is necessary and when it is not, if people are in sustained conflict of non-moral interests that they cannot effectively resolve by appeal to their moral codes or norms, then, for the purposes of social order, they are in a moral vacuum. “Traditional” morality is no help, but neither is it the driver of the conflict. I think it would be plausible to tell a history in which, from the tenth to the sixteenth centuries, many in northern Europe developed moral perspectives that radically loosened the claim of morality over trading and employment, thus allowing tremendous latitude for social organization via self-interested bargains. As Gauthier (1986: chap. 4) observed, the market became something of a morality-free zone. Even here we don’t want to follow some of my economist friends and make the mistake of thinking that markets simply substitute prudence for moral regulation: we now have a good body of evidence that without effective moral norms, markets can be grossly dysfunctional (Rose, 2011; Zak, 2008;

Friedman, 2008). Perhaps a better example is the so-called “realist” view of international relations as a sphere outside of morality, in which states are guided solely by national interest. Again, I think this was seldom if ever purely the case — even realists such as Hans Morgenthau (1973: 230) ultimately denied it. Yet there is here an important insight about the place of mutual benefit and instrumental rationality when morality is of no help. Notice, though, that this is not really a case of deep moral *conflict*: morality is not employed to resolve conflicts, but neither is it generating them.

I am more skeptical that the reasoning of *homo prudens* is of much use in cases characterized by The Strong Centrifugal Condition. Here it is not simply that moral difference undermines morality’s ability to bring us together: morality is the force driving us apart. And pointing out that moral conflict is imprudent often is beside the point. Many people think a moral person is committed to imprudent actions: running the risk of serious conflict and the breakdown of critical features of cooperation is precisely what their moral commitments require.<sup>2</sup> In such cases, the choices of *homo prudens* are not ones with which agents identify. I stress that this does not show that Moehler’s argument is invalid, for he explicitly restricts it to cases where peace is preferred to further conflict (102). The point, rather, is that the model is not of obvious relevance in societies riven by deep moral conflict (i.e., when the Strong Centrifugal Condition holds).

#### 4 How Deep a Conflict?

So my worry is this: Minimal Morality presents itself as a solution to moral conflict that is so deep and divisive that “traditional” moral justification of the rules of cooperation is well-nigh impossible, but it is precisely in such contexts that morally committed individuals are apt to reject *homo prudens* as a model for conflict resolution. Think of England in 1640, France in 1789, America in 1861 or contemporary Islamic societies split between Shia and Sunni, or India between Hindus and Muslims. And the list goes on and on. To different degrees, on moral, ideological or religious grounds, large groups act resolutely non-prudentially: they are willing to set back their interests and incur deeply impaired social cooperation and risk

---

<sup>2</sup> If commitment to moral principles is prominent in states’ foreign policy, Morgenthau (1973: 252-6) insisted, we are apt to have a highly conflictual international system — which is why he thought moralism in foreign affairs was so dangerous. It would override prudential pursuit of national interests.

various levels of violence in the name of fidelity to moral or transcendental ideals. In these cases, I do not see how effective social cooperation can be grounded on prudential morality.

A familiar stylized history of the growth of liberal toleration — from a *modus vivendi* to liberal justice — seems to support a prudence-first analysis. As the tale goes, after the wars of religion, the parties were exhausted and agreed to call off the conflict for purely prudential reasons. Eventually this self-interested settlement evolved into a moral consensus on liberal toleration. There is certainly some truth in this tale: people can come to conclude that their moral battles are simply more destructive than they are worth. Yet, I reckon this was a far more complex process than this simple narrative supposes. For *as* the conflict progressed, many came to reevaluate whether their religious beliefs really called for such extreme disorder. The English Civil War lasted from, say, 1642-51. Hobbes published *Leviathan* in 1651, giving the *homo prudens* case for civil order — which fits nicely into the familiar prudence-first narrative. But Locke's *Letter Concerning Toleration* was published only forty years later in 1689, and there Locke argues that properly understood Christianity calls for toleration. Running in tandem with the prudential case for peace was a fundamental reevaluation of the religious grounds for conflict. Rather than the prudential simply overriding the religious, what emerged by the end of the seventeenth century seemed more like a new congruence between prudential interest and religious belief.

##### 5. A Congruence Problem for Minimal Morality

A morality based on the value of peace and prudence confronts the mirror-image of Rawls's congruence problem. In *Theory* (1999: 496-99) Rawls derived principles of justice from our common commitments, but then worried whether the demands of justice might be overridden by a person's reasons emanating from self-interest or, more generally, their conception of the good. Thus he sought to show congruence between the right and the good. In parts of the *Political Liberalism* project this led to the claim that the principles of justice were not *fully justified* until they were justified by each person's entire conception of the good (Gaus and Van Schoelandt, 2017). Now Minimal Morality's problem is basically the opposite: having derived morality from *homo prudens* — a model of agency constrained by minimalist shared values — the question arises whether the results remain justified in light of a person's moral commitments. "Like Kant's moral theory," Moehler writes, "for the domain of pure instrumental morality," his "contractarian theory demands that agents *abstract* from their

particular interests and focus on the general needs and interests of all human beings as separate autonomous beings. In doing so, it demands that agents take on a multiperspectival standpoint" (133, emphasis added). Thus Moehler faces Rawls's full justification worry: does the justification of Minimal Morality survive once this abstraction is relaxed? For Minimal Morality to be relevant people must be able to affirm it, and its high valuing of peace, in light of their total commitments. If not, then while the derivation of Minimal Morality may well be valid (as, Moehler (30) points out, Rawls's derivation of justice as fairness in the original position may be) the justification would be overridden by other values of the parties. Just as a citizen may press Rawls: "Why should I care what would be chosen in the original position, given that it draws on only some of my values?" they can press Moehler, "Why I should identify with the choices of *homo prudens* in the empathetic contactor situation if its conflict-aversion is inconsistent with my convictions about the demands of the holy or just?" This is a general problem with the all versions of the Shared Basis Approach: if we are to identify with the conclusions of those who reason only on the basis of shared reasons, we have to know why *their* reasoning shows *us* what our moral reasons are (see Thrasher, 2017). Note that when The Weak Centrifugal Condition obtains, this is relatively easy: morality is not really relevant to the dispute, so being prudent seems pretty unobjectionable, and we may all affirm Minimal Morality. But it is a lot harder in conflicts characterized by The Strong Centrifugal Condition.

#### 6. *Is One's Morality a Bargaining Chip?*

It is important to stress that *homo prudens* is not necessarily self-interested. "The *homo prudens* model," Moehler tells us, "considers all types of interest that agents may have, independent of their motivation and content" (102). So suppose we take all of a person's concerns — and this certainly can include her ideals and religious commitments — and devise a preference ordering. What is important is that this ordering is constrained by the *homo prudens* assumptions: people value peace over conflict, and demand that their autonomy is respected and basic needs met.

Moehler describes all these values and commitments that underlie a person's utility function as her "interests," so (speaking roughly now) maximizing utility just means that she maximizes the satisfaction of her cardinally represented interests. Once prudent individuals are represented by such well-formed utility functions, the stage is set for Moehler to employ



a version of the Nash bargaining solution. This is an influential proposal today; Ryan Muldoon (2016: 79-80) employs the Nash bargaining solution in a related way. Many hold the Nash solution to provide a unique, rational, bargain between morally diverse individuals. As much as I respect Michael and Ryan, I remain unconvinced that this particular formal solution enlightens us about our moral disputes.

My worry is that the prudential individual's utility function is based not simply on things that we might rationally haggle over — say our respective shares of a common product — but about our moral commitments. If we do not keep in mind that some of the “preferential” differences represent disagreements about morality and justice, it may seem when Alf and Betty have a moral conflict we can usefully model this in the same way as a problem about dividing a resource (see Gaus, 2018a). Suppose Alf believes that God instructs that living with the heterodox endangers the salvation of all, while Betty advocates toleration on Lockean grounds. The question is this: can either Alf or Betty accept the idea that the moral resolution of this dispute is defined by the stabilized Nash bargaining solution? “Come on Alf,” she says, “be reasonable and meet me half-way: let's say we shall have toleration except for Catholics and — to concede something to you — we shall also exclude Methodists.” Could Alf really think it would be unreasonable for him to resist this, in the same way as it would be unreasonable for him to insist, say, on all the cake they have just made? James II essentially offered toleration for Dissenters if they would endorse tolerance for Catholics — it was an offer they found reprehensible and a violation of their religious convictions. I am not saying that parties might not, at some point, end up in some sort of pragmatic compromise, but I cannot see that they would see the compromise as *any* sort of a morality. Once everything is turned into an “interest” it is too easy to conflate the sort of haggling consistent with moral commitment and that which is not. If we are apt to come to blows over conflicts of our cherished ideals, instructing us to be prudent and bargain does not look reasonable.

### *7. Reconciliation v. Splitting the Difference*

I am an enthusiastic advocate of moral reconciliation on matters basic to social cooperation — of each finding, within his moral view, the resources to live on moral terms with others even if they are not the terms that he sees as morally best. But, like Rawls's overlapping consensus, this must be a task for each moral view to work out on its own, as it seeks to live with others on moral terms. No bargaining theory is going to give a universal answer: John

Nash, alas, did not discover a single rational recipe for moral reconciliation. There is no recipe, simply because one dimension of moral conflict is the importance of reconciliation (Gaus, 2018a). As I noted at the outset, this is a feature of the Diverse Moral Inputs Approach that frustrates many philosophers: there is no canonical derivation or decision method that gives the unique answer. It is always a matter of each moral agent, in light of her and others' moral views, sometimes adjusting her moral demands, seeking partners in moral relations, gravitating to different moral rules in different contexts, and forming moral orders as they do so. This is a moral theory of dynamics, not statics: we look for ways of relating that facilitate maximal reconciliation, not generate recipes for it.

#### *8. Are Prudence and Rational Choice Really Fundamental?*

So, I'm afraid that after wrestling with Hobbesian-inspired theory for most of my career, I remain unconvinced that, for all its elegance and sophistication, Hobbesian-inspired rational choice theory tells us much about the how to live in morally divided societies. I wish it did, for I would love to have on call on a minimal, back-up, theory, which could assure us that there is almost always some specific rational morality on which we can draw. Moehler's analysis, is in the end a hopeful one. As long as we are prudent, there is at least a Minimal Morality for us: if we go far down enough, below our moral disputes, we share a commitment to prudential reason. Perhaps my fundamental disagreement with Hobbesian analysis is that, on my reading of human evolution and cognition, moral structure and rules come before social order by rational choice (Gaus, 2018b, 2018c). We are not rational creatures who invented social morality: we are social creatures who hit upon morality, which then allowed us room to further structure social life through mutual advantage. As Joseph Henrich (2016: 51-53) observes, strategic reason is only useful given a set of shared norms: it is neither prior to them, nor can it replace them. When we truly have no shared "traditional" morality to guide us, I cannot see much possibility of fruitful and reasonably stable social cooperation.

## WORKS CITED

- Friedman, D. 2008. *Morals and Markets*. New York: Palgrave Macmillan.
- Gaus, G. 2011. *The Order of Public Reason*. New York: Cambridge University Press.
- Gaus, G. 2018a. "Self-organizing Moral Systems: Beyond Social Contract Theory." *Politics, Philosophy and Economics*, vol. 17 (May 2018): 119-147.
- Gaus, G. 2018b. "The Priority of Social Morality." In *Morality, Governance, and Social Institutions: Reflections on Russell Hardin*, edited by Thomas Christiano, Ingrid Creppell and Jack Knight. New York: Palgrave Macmillan, 2018: 23-57.
- Gaus, G. 2018c. "It Can't Be Rational Choice All the Way Down: Comprehensive Hobbesianism and the Origins of the Moral Order." In *Tensions in the Political Economy Project of James M. Buchanan*, edited by Peter J. Boettke, Virgil Henry Storr, and Solomon Stein. Arlington, VA: Mercatus Center, 2018.
- Gaus, G. and C. Van Schoelandt. "Consensus on What? Convergence for What? Four Models of Political Liberalism." *Ethics*, vol. 128 (October 2017): 145-172.
- Gauthier, D. 1986. *Morals by Agreement*. Oxford: Clarendon Press.
- Henrich, J. 2016. *The Secret of Our Success*. Princeton: Princeton University Press.
- Kogelmann, B. 2017. *Agreement, All the Way Up: An Essay on Public Reason and Theory Choice*. Doctoral dissertation, University of Arizona.
- Macpherson, J. M. 1988. *The Battle Cry of Freedom*. New York: Oxford University Press.
- Moehler, M. *Minimal Morality*. New York: Oxford University Press
- Morgenthau, H. J. 1973. *Politics Among Nations*, fifth edn. New York: Alfred A. Knopf.
- Muldoon, R. 2016. *Social Contract Theory for a Diverse World*. New York: Routledge.
- Rawls, J. 1999. *A Theory of Justice*, revised edn. Cambridge, MA: Belknap Press of Harvard University Press.
- Rawls, J. 2005. *Political Liberalism*, expanded edition. New York: Columbia University Press.
- Rose, D. C. 2011. *The Moral Foundations of Economic Behavior*. New York: Oxford University Press.
- Thrasher, J. 2017. "Constructivism, Representation, and Stability: Path-Dependence in Public Reason Theories of Justice." *Synthese*, DOI 10.1007/s11229-017-1488-7.
- Tomasi, J. 2012. *Free Market Fairness*. Princeton: Princeton University Press.
- Vallier, K. 2014. *Liberal Politics and Public Faith*. New York: Routledge.
- Zak, P. J. (ed.). 2008 *Moral Markets*. Princeton University Press.