The Good, the Bad, and the Ugly:

Three Agent-type Challenges to *The Order of Public Reason*[*]

Gerald Gaus

1 Agent-type Challenges

Richard Arneson, Jonathan Quong and Robert Talisse are three of today's leading political philosophers. To have them carefully attend to my book is most gratifying; to have the opportunity to reflect on their concerns is a real honor. Their essays are fair-minded, well-focused and, of course, insightful and probing (which, alas, is not always the case with critics), so it is truly enlightening to reflect on them. It is also surprisingly easy to organize my comments, for in different ways, all press what I shall call "agent-type challenges" to the project of *The Order of Public Reason* (*OPR*). In different ways they all focus on a type (or types) of moral (or sometimes not-so-moral) agent, and object that *OPR* does not give a satisfying account of what we are to say to, or how we should act towards, such agents. In my view these are surprisingly difficult objections to analyze, since, as we shall see, they often turn on details of the description of the relevant agents. But let us see what progress we can make here. I hope to show that *OPR* does not say quite what they think it says, and it often leaves them room for saying what they would like to say about such agents.

2. The Background: Self-sustaining Moral Constitutions

Arneson, Quong and Talisse provide very nice and fair summaries. However, let me begin by stressing a few fundamental features of the analysis that I think we have to keep firmly in mind.

Most importantly, *The Order of Public Reason* is not a work in moral philosophy as that is normally understood. In two important, and I think largely misunderstood, essays from the 1970s, Rawls sought to characterize a distinct form of moral inquiry that he deemed "moral theory," which is not concerned with uncovering the "moral truth" but, rather, is a

> …search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society…. The task is to articulate a public conception of justice *that all can live with*…. What justifies a conception of justice is not its being true to an antecedent order to and given to us, but its congruence with our deepest understanding of ourselves and our aspirations, and our realization that, given *our* history and the traditions embedded in *our* public life, it is the most reasonable doctrine *for us*. *We can find no better basic charter for our social world*.[1]

Moral theory thus understood is a search for what Rawls calls a "moral constitution."[2] Rawls never claims that moral theory's inquiry into this moral constitution *supplants* traditional moral philosophy and religion understood as the search for moral truth; rather is *supplements* them in the sense that it seeks to arrive at a shared moral framework *all can live with* in a social world where our understandings of moral truth clash. Moral theory is a search for a "constitution" (in the first instance moral, not political) that allows public adjudication of diverse individual claims based on differing moral and valuational perspectives.

The question arises: why should we be interested in such a constitution, such a shared moral framework that we all can live with? Each of has her own deep convictions about the moral truth and normativity, the nature of our place in the universe, and our ideal understandings of human relations. Why do we need a basic social-moral framework that is apt to depart from these? We can give both a subjective and an objective answer to this crucial query.[3]

From our subjective, participant, experience, we find ourselves committed to a web of moral practices and reactive attitudes. We expect certain actions as things we can demand of others, even strangers: we experience resentment and indignation when we see that we or others are treated in violation of our shared morality; we blame, we criticize, we punish. We hold others responsible, and insist that they knew better than to act in those ways. Our values and ways of living, our understanding of our relations with our fellows and our ability to trust others, are all informed by this way of seeing of the world. As P. F. Strawson taught us, as participants in the moral enterprise, this is the view we are driven to take, and it informs how we see ourselves and others.[4] But Strawson also taught us that we can step back, and look at the moral enterprise from an objective point of view. From this objective point of view, the regulation of social relations through this sort of moral practice is one of the foundations of human social life. No human society has been able to exist without one; societies with weak or ineffective frameworks are characterized by serious social dysfunctions such as lack of trust between strangers and endemic conflict. A shared moral framework is perhaps the distinctive human achievement, which allows a deeply social and cooperative existence among creatures who are not related by kinship or a common view of the world in which they live.[5]

So *OPR*'s focus is this social-moral framework that sustains human cooperative life — something that has been more or less invisible in late twentieth/early twenty-first century normative ethics and social philosophy with its obsession with defending highly controversial accounts of TRUE morality and justice. Like a political constitution, a modern moral constitution is a structure we all can live with though we deeply disagree on ideals. It will do no good to propose a constitution that is a litany of one's cherished ideals; under modern conditions we need a constitution *just because* we share a deeply social life with those who do not share our cherished ideals — our fundamental convictions that we have THE TRUTH on fundamental issues. Liberalism only arose in the sixteenth and seventeenth centuries when disparate religious groups came to converge on this conception of the political constitution.

What we need is a moral framework that supports social cooperation and sustains the reactive attitudes of the participants; it must ground those practices of resentment, indignation and blame that inform our participant view of moral life. And, crucially, the requirements of the framework must be internalized by its participants; they must see the rules of the constitution as things they care about, which, from their reflective perspectives, they can endorse because it coheres with their evaluative commitments and standards. Only if it does this will the constitution exhibit internal stability or robustness.[6] There is great dispute about the place of stability in Rawls's system, but once we appreciate that the aim of moral *theory* is a moral constitution that adjudicates claims among those with deeply different beliefs about the moral truth in a way that all can live with (given their own cherished normative commitments), we can see that there is no point to a moral constitutional structure that does not have strong tendencies to maintain itself.

Structures of cooperation that do not generate their own support are ineffective and short-lived.

And so we come to the crucial point: how are we to have such a constitution? One possibility is that committed zealots, relying on political and social power and using indoctrination and severe punishments, get the rest of us to act as they are convinced in their moral or religious consciences we should. But even if this can achieve a sort of stability, it cannot sustain our practice of social morality. Unless participants have some sort of access to the reasons behind moral demands, their failure to comply will not invoke rational reactive attitudes. Our rulers might be angered or frustrated by our failure to do as we are told, but they cannot say that we knew better, or could reasonably have been expected to know better, or that our failure to confirm manifested ill-will. If someone truly cannot reason his way to seeing why he should ϕ, I cannot read into his failure to ϕ any ill-will towards me or others. Assuming his "ignorance" is not itself culpable, I am confronted with someone who cannot get why he should do as I say. Ignorant, perhaps; ill-willed, no.

We are thus led to the public reason project: how can a self-sustaining moral constitution exist among free individuals, who think for themselves? This is not an esoteric commitment of wayward Kantians, and it certainly is not a project in moral philosophy, seeking to weigh in with its own, controversial, doctrine of the sources of normativity. Rather, the public reason project investigates whether a self-sustaining moral constitution is possible among deeply diverse individuals, most of whom are confident in their insight into the nature of moral and religious truth. *The Order of Public Reason* is thus best understood as a model of how a free self-sustaining moral constitution can be extended as far as possible. How inclusive can such a moral framework be? We live in a world with deep diversity, and for many

modern societies — for example European societies that now have significant Muslim populations — are facing a higher degree of diversity than has occurred in centuries. The question confronting us is how extensive and diverse a community can be and yet sustain a moral constitution that all endorse.

3. The Good

The first challenge to this moral constitutional project is to inquire whether a good and moral person could possibly endorse it. Suppose, following Arneson, social morality contains a rule *R*, which requires *X* in current circumstances.

> This social morality now dictates that Allessandra should do *X*. However, Allessandra judges that according to her own moral views, refraining from doing *X* would be the morally superior course of action. In arriving at this decision, Allessandra gives proper weight, from her own standpoint, to the indirect moral costs of her proposed course that would stem from such possibilities as that others will witness her violation of the going morality and become less firmly disposed to conform to it across the board. However, Allessandra has good reason to believe her judgment in this matter, though fallible, is superior to the judgment of those who uphold the conventional morality on this issue. (Arneson ms. p 5)

So we suppose that Allessandra believes that on the best view of morality she should not-*X*; given her normative perspective, which she believes to be sound, not-*X* is ranked as a better action than *X*-ing. Importantly, Arneson stipulates that Allessandra accepts that rule *R*, which requires action *X* in the current circumstances, is what *OPR* calls an "eligible" rule. So on her normative perspective she

acknowledges significant moral reasons to endorse a social world in which all internalize *R*, feel guilt for violating *R* (i.e., for not *X*-ing in the current circumstances). And, crucially, given that *R* is seen as an eligible rule, Allessandra's normative perspective ranks this social world, in which all internalize, act on, and feel guilt for violating, *R* as morally better than a moral world where there is no rule of social morality. Her ordering thus comes out to:

(*a*) *R* (a rule of social morality) requires *X*, but I not-*X*

(*b*) *R* requires *X*, and I *X*.

(*c*) There is no rule of social morality about this matter.

By Arneson's stipulation she is committed to $a \succ b$; by the definition of an eligible rule she is committed to $b \succ c$.[7] But note that we can further specify the state of affairs regarding (*c*) by dividing it into:

(c_1) There is no rule of social morality and I not-*X*.

(c_2) There is no rule of social morality and I *X*.

Given Allessandra's stipulated moral view, I suppose $c_1 \succ c_2$. Now notice that she holds $a \succ c_1$; she ranks as higher the state of affairs in which *others* act on and internalize *R*, feel guilt and so on but *she does not*, to the social world where there is no rule of social morality and each simply follows her own individual conscience. So she wants *others* to accept *R* as normative and conform, but for her (and perhaps those who agree with her) to be exempt (not of course officially, for the rule that gave her official exemption almost surely would not be endorsed by others). Presumably she wants others to internalize and obey *R* because a socially accepted rule has great social and moral benefits (coordinating our activities, providing shared expectations about when moral criticism will be forthcoming, solidifying trust, regularizing punishment and so on) and *R* actually sufficiently coheres with

her normative views, but she has, from her perspective, moral reasons for defecting and free-riding. And free-riding she is, for she endorses the moral rule for all others ($a \succ c_1$), but wants the moral benefits of R without her paying the moral costs of her complying ($a \succ b$). Arneson announces that he sides with Allessandra; he agrees with her that acting on R "lacks supporting moral reasons that should give it a claim to our allegiance." (Arneson ms. p. 6). It thus seems to follow that on Arneson's analysis that a good moral person ought to free-ride on a community that acts on moral rules — rules she endorses from her normative perspective as something for others to act upon. She must, as it were, be an individual moral optimizer though she endorses a social world where the others are not.

I must confess I do not find Allessandra such an admirable type after all: she endorses a social world where the rest of us follow the moral constitution, feel guilt for violating its rules and so on, but as she is special, she defects when doing so better conforms to her personal judgments about morality.[8] In any event, on the analysis of *OPR* Allessandra does not truly endorse R as an eligible rule, for a necessary condition for doing so is that she endorses it when it applies to *her*, regardless of the position she occupies (this is the reversibility requirement; *OPR*, 299-301). Given $a \succ b$, she prefers violating the rule when others do to conforming to it. She does not internalize the rule herself, though she wants others to do so. In the terms of *OPR*, that ordering shows that her endorsement of R is not in equilibrium; she does not truly endorse R as an eligible rule.

An Allessandra who does not take this route of secret defection, would have to have the following ordering:

(*d*) Rule R^*, which requires not-X in these current circumstances.

(*e*) No rule of social morality (and because Allessandra is under no requirement by the moral constitution to *X*, she not-*X*s).

(*f*) Rule *R,* requiring *X* in these current circumstances.

Here we can suppose that her highest ranked option is that we secure all the benefits of a moral constitution that has a common moral rule about this matter (coordinating our activities through internalized moral guidance, providing shared expectations about when moral criticism will be forthcoming, solidifying trust, regularizing moral punishment and so on) so long as, from Allessandra's perspective, we have the optimal moral rule. Here she gets all the moral benefits of a shared moral framework and incurs zero moral costs (always a happy a result). But, as a relentless personal moral optimizer, as soon as a rule in not the best in her eyes, she rejects it. She ranks as ineligible all non-optimific rules; she refuses to perform *X* in these circumstances, and so can acknowledge no authoritative rule that she must *X*, thus ranking as better a social world with no shared moral rule, where she violates no requirement by failing to *X*. Thus *d≻e≻f*.

This second version of Allessandra is much more of a worry; indeed it is a view implicit in much moral philosophy. To many moral philosophers it is incoherent to admit a moral requirement that, from one's moral point of view, does not instruct us to do the morally best action. But now we come to the core question: does Allessandra truly optimize by this ordering? In her first incarnation above, she endorsed everyone *else* internalizing the rule, apparently because (as, say, a consequentialist) she recognized the great moral benefits of a shared, moralized, social life. But once we have rejected the free-riding strategy, Allessandra will see that taking up her optimizing stance precludes achieving those moral goods. In a world of moral disagreement, a society of such optimizers will not share a moral

constitution, for each person's eligible set is restricted to the constitutional rule that she considers best, and most others do not share one's view of the morally best. So this version of Allessandra-type optimizers will surely end up in a social world in which there are no shared moral rules. Instead, we will have a social world in which each permissibly acts on her own moral view as she sees fit. This was precisely Kant's understanding of the state of nature, where each claims "the right to do what *seems just and good to him*, entirely independently of the opinion of others."[9] Kant was struck by the conflict that he thought would be endemic to such a situation, but suppose that we admit that law might restrain people from overt conflict. We would still be left bereft of moral goods that all societies require, and that all participants in our moral life presuppose: a settled basis for the reactive attitudes, common understandings of the appropriateness of moral condemnation, and a common understanding of when one can issue moral demands that will have uptake by other competent moral agents. We are left either with morality banished to the private sphere (each has her own personal standard of what should be done), or a jostling in public of opposed voices, hectoring each other, but having few expectations of the terms of their moral interactions. As the social contract tradition has always stressed, once we realize that social life — and that includes moral life — is a social, cooperative, endeavor, narrow personal maximization is almost always self-defeating strategy.[10]

None of this supposes that Allessandra should renounce her judgments of what is morally the best, or stop trying to change the views of others. Moral reform is horribly under-theorized in contemporary moral philosophy; the brunt of effort is investigating types of truth claims and different theories of the best, as if moral philosophy was a scientific discourse where once we arrive at a true claim, its dissemination among competent inquirers will pretty much take care of itself. The

good have great and important work to do, but that is neither to haughtily exempt themselves from our shared moral life or undermine it because it does not meet their personal convictions about the best.

4. The Bad

Quong's type-objection focuses on Carl, who has "a sufficient reason to believe, for example, that it is morally required for him to kill infidels (call this belief KI)" (Quong ms. p. 4). Quong raises the spectre of Carl because, on the justificatory model of *OPR*, we start out with a basic conception of a boundedly rational agent, one based on the rather austere model of rational choice theory. Like Rawls — and unlike Quong — the aim is to derive moral conclusions from an essentially procedural model of the rational *plus* substantive commitments that we model into the agents that are not part and parcel of being rational (for Rawls, this is the reasonable). Neither Rawls nor I wish to build moral conclusions into the very idea of the rational (what Quong calls a "substantive" conception of the rational): on this substantive view a person is rational only if she recognizes certain substantive reasons to act. Now I entirely agree with Quong that if we build into the rational that a rational person must acknowledge a moral reason to *X*, then we can easily (and rather too quickly) show that all rational people have reason to *X*. We could do then what Rawls says that we cannot: deduce at least some of morality from the concept of the rational, since we have built some of it into the rational. *Pace* Quong, Rawls insists that "[w]hat rational agents lack is the particular form of moral sensibility that underlies the desire to engage in fair cooperation as such, and to do so on terms that others as equals might reasonably be expected to endorse."[11] The public reason project, as Rawls and I understand it, is to commence with a non-substantive conception of rational agents, and then to see what substantive concerns (concerns

that are not part and parcel of the rational, but are broadly within "the reasonable") would lead agents to endorse a shared framework.

OPR thus introduces a number of substantive concerns that we attribute to fellow moral agents that go beyond what makes a person rational. I suppose, as I have said, that each of us (including Allessandra) is centrally interested in sharing a social moral life with those who are good-willed, appreciate the benefits of a social existence, and are employing evaluative considerations that strike each other as intelligible and relevant to the question at hand, and have employed their reasoning on these commitments to a required degree. Later I introduce additional assumptions about their conceptions of themselves as agents. All of these are aspects of our moral sensibility; the folks whom you and I see as the population with whom we could well share a moral constitution, if only we could find one. By separating these commitments from the concept of the rational, like Rawls, *OPR* then constructs a model to ask: are these sufficient constraints such that those who meet them are apt to have reason to endorse a moral constitution? If not, then we may have to go back and make more restrictive assumptions about the sort of agents that can share a social moral framework. But given the approach, we absolutely must keep distinct the rational and the reasonable — what it is to be merely rational, and what additional sensibilities and concerns are needed to render one a participant in the moral constitution of a society. We do not wish to build our moral conclusions into our concept of the rational.

Given this, Quong is entirely correct that it must be possible for a rational person to endorse the belief that he ought to kill infidels. This, indeed, is axiomatic: because the rational does not include moral sensibility and a concern for a cooperative social life (and from the details given, I would say that Carl is a bit lacking on the good will

dimension), we can describe a rational agent who lacks the moral commitments to endorse a moral constitution. So we must allow:

(1) Carl is a conceptually possible type of agent.

It is, however, unclear how much more can be said without adding more details to the Carl story. To fix a case that seems maximally friendly to Quong's challenge, let us stipulate that, say, Carl's country is not at war with the infidels; rather, he lives in the same society as those whom he would kill. Let us further suppose that Carl sees himself as under a religious obligation to rid his society of these people. Further stipulate that Carl is boundedly rational, and given as much deliberation as we would expect from a normal moral agents, he endorses the belief "KI." I admit all this as a stipulation of a sort of conceptually possible case; I believe that it is anything but common.[12]

As Talisse correctly points out, much of the analysis rests on the "Basic Principle of Public Justification:"

A moral imperative "ϕ!" in context C, based on a rule L, is an authoritative requirement of social morality only if each normal moral agent has sufficient reason to (*a*) internalize rule L, (*b*) hold that L requires Φ–type acts in C, and (*c*) moral agents generally conform to L.

Because (by stipulation) Carl does not have sufficient reason to endorse a rule against killing those who are infidels, the general rule against killing is not justified to Carl, and so has no authority over him. Thus we must also concur with Quong that:

(2) The moral constitution of Carl's society, and in particular the general rule against killing, has no moral authority over Carl.

So far, this is all straightforward. Things get a wee bit more complicated when we analyze what it means to say that a rule has moral authority, and the reasoning that leads up to the Basic Principle of Public Justification. In many ways this is the heart of the book, so I shall have to content myself here with summarizing the main points.

(*i*) *OPR*'s core problem is the interpersonal authority relation in which Alf morally demands that Betty perform, or refrain from, some action ϕ. Betty is about to take Alf's property; Alf issues a demand that she not do so. If she fails to comply, he will experience the reactive attitudes, morally criticize her, and he holds that she should feel guilt and remorse.

(*ii*) For Alf to possess justified moral authority over Betty, both an empirical and a normative condition must be met:

(a) The *empirical condition* is that there is some socially agreed-on trait or feature of their interaction such that Betty can confirm "Yes, this is a situation in which Alf holds authority." Unless this empirical condition is met, moral authority cannot serve its function of adjudicating conflicting claims and coordinating expectations. In the context of *a moral theory*, seeking to investigate the conditions for an effective self-sustaining moral framework, this empirical condition is critical. An "authority relation" that did not meet the empirical condition would serve no practical purpose. As T.H. Green remarked in the context of political authority, appealing an "authority relation" that had no empirical recognition is like appeal to "the mere name of a fallen dynasty exercising no control over men in their dealings with each other."[13] On some hierarchical views of moral life, the empirical condition is some mark that Alf but not Betty possesses (e.g., being a male, having a PhD

in Philosophy). The task of *OPR* is to develop an egalitarian theory of moral authority; in *OPR* the empirical indication that Alf possess authority is that this is a situation *C*, in which Rule *R* applies, and the rule grants Alf authority to demand φ.

(b) The *normative condition* is that this rule *R*, which grounds the claim to interpersonal authority, is one that both Alf and Betty, reasoning as well as our practice of morality requires them to reason, would both endorse. In the framework of *OPR* a claim to authority that does not meet this normative condition is *authoritarian*: Alf is seeking to morally command Betty's action to act in ways that, given her normative perspective, she cannot see as morally required.

(*iii*) The justifications for this egalitarian normative condition are two:

(a) If we consider a commitment *exogenous* to our practice of morality — that is, an external requirement that we bring to bear in evaluating our practice — the normative condition expresses the idea that we are to treat all as free and equal interpreters of the moral law. That is the conception of free and equal personhood at the heart of *OPR*. In a world of deep disagreement about what morality requires, none have the mark of rightful authority such that they are endowed with a justified authority to command others as to what their moral responsibilities are.

(b) An important claim of *OPR* is that this exogenous egalitarian standard is also the standard of justified authority needed to maintain the reactive attitudes at the heart of social morality. Thus there is a parallel *endogenous* argument: only when Betty can reason herself into endorsing *R* as a moral

rule can Alf rationally sustain the reactive attitudes if she fails to φ. If she cannot reason herself into endorsing *R*, Alf cannot rationally view her failure to φ as a manifestation of ill-will. So to reject the egalitarian normative condition is to undermine our current deeply-held understanding of our moral relations to others, and what is involved in sharing a moral life with others.

Given this sketch, it is now clear why Betty cannot claim authority over Carl; the normative condition (*ii*b) is not met. Thus any authority that Betty would claim over Carl would have to be an inegalitarian authority — an authoritarianism ("liberal authoritarianism" is not, after all, an oxymoron). Now Betty, with whom Quong sides, insists on the authority of the *R*.[14] Betty says to Carl: "I'm asserting that there are decisive reasons for you not to kill infidels, and even if there is a procedure by which this substantive claim can be rejected, this *does not impugn the normative authority of the substantive claim*" (emphasis added, Quong ms. p.7).

Note that Betty tries not to talk about authority as an interpersonal relation or a social device by which we resolve disputes. It is, she says, not a question of *her* authority, but the authority of the "substantive reason" to which Carl must submit. And we get to the crux of my disagreement with Quong (and, I must admit, many others who talk this way). Betty seems to be claiming that somehow the substantive reason simply speaks *through* her, as if she was an oracle, allowing true normativity or real substantive reasons to use her speech to instruct us. (I cannot help but picturing a séance). But as Hobbes so eloquently argued, though each person is convinced that she is reporting right reason, these reports are always mediated through her reason, and our dispute is about whose use of reason has gone astray. If our fundamental disputes are about who has reasoned themselves to the correct

moral conclusion, seeking to resolve those disputes by saying that reason has the true authority, and you speak for it, simply fails to recognize the nature of the dispute. In a passage that is the fountainhead of public reason political theory (and which, alas, I admit I have quoted too often) Hobbes observes: "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 their want of right reason, by the claim they lay to it." [15]

> Now it is absolutely essential to realize that:
>
> (3) Proposition 2 above does not imply that it is impermissible for others in Carl's society (e.g., those who do endorse the moral constitution) to enforce their moral requirements and make Carl ϕ, nor does proposition 2 imply that it is impermissible to legally punish him for failure to ϕ. [16]

Proposition 3 merely points out that the Basic Principle of Public Justification specifies conditions for moral authority, not moral permissibility; failing to satisfy the Basic Principle leads to proposition 2, but not to any claim about impermissibility of actions. Unless the story is changed, it is certainly permissible, and indeed praiseworthy, to stop him from killing.[17] What we cannot do is to suppose that we can issue authoritative moral demands to him; when we enforce our moral constitution against Carl we do not treat him as a free and equal moral reasoner. Given that the normative condition (*ii*b) is not satisfied, we can no longer treat him as a fully responsible moral agent. The reactive attitudes are undermined. But *in extremis* we must act as we must. As Adam Smith observed about such cases,

"a good man will always punish them with reluctance, when they evidently proceed from false notions of religious duty. He will never feel against those who commit them that indignation which he feels against other criminals, but will rather regret, and sometimes even admire their unfortunate firmness and magnanimity, at the very time that he punishes their crime."[18]

A final word about human rights. It is crucial to stress that Carl is indeed an extreme case, not a normal instance of appeal to human rights. In the sort of conditions we have been analyzing — where the party being imposed upon does not recognize the authority of the claims made upon him — human rights-based interventions are apt to be costly and often ineffective. Unless the enforcers have overwhelming force, their intervention is likely to simply induce counter-force. Human rights are not well protected by liberal egalitarians imposing them on the rest of the world on the grounds that *they* truly know the "substantive reasons" there are, though the imposed-upon cannot reason themselves into seeing them. Human rights are made real by the development of global institutions that connect well-defined human rights claims to diverse social and cultural practices, and so are not seen as impositions of liberal authoritarians. In an effective regime of human rights, they are not merely convictions of egalitarian liberals about what those others must do, but are embedded in widely-embraced institutional structures that are recognized as authoritative — they are human rights in an order of public reason.

5. The Ugly

A central claim of *OPR* is that the rules of a justified social morality provide standing for a member of the public to demand compliance by an offender. It is, as I say, everyone's business that others conform to the moral constitution, and so everyone

has standing to insist that others comply. And so we come to our third agent-type challenge, the rather morally unattractive Reverend Wholesome, who

> routinely publicly chastises adulterers and regularly demands of others that they "stay faithful!" Once the Reverend is revealed to be a serial adulterer, it is perfectly reasonable for others to reject the authority of *his* demands concerning adultery, even though the content of that demand is indeed authoritative. The Reverend's hypocrisy serves to revoke his standing to speak for social morality, even when he demands what it in fact requires. (Talisse ms. p. 13).

Talisse suggests that joining the morally ugly Reverend Wholesome are "others who are morally lost, benighted, and blind." While good liberals are "fully prepared to live with those they regard as morally benighted on fair and cooperative terms," many are convinced that "they are conscience-bound to reject the thought that their benighted fellows could be fit to wield *moral* authority of any kind." (Talisse ms. p. 14).

I am a little unsure of whether this is an ideal case to consider. For one, today many doubt whether adultery stills runs afoul of general social morality; while some continue to see it a violation of our shared morality, to many marriage is a private contract, and its terms are open to ongoing renegotiation by the parties involved. In this case we would think whether one engages in extra-marital sexual relations really isn't other people's business. And even if it is still part of social morality, marriage relations are themselves protected by rights to privacy — an additional rule that provides them with special protection from scrutiny.[19] In addition, it is somewhat unfortunate that, as with Quong, the target is once again a religious person, someone with whom philosophers are not apt to feel much sympathy in the first place.[20] So we have a norm that most philosophers are suspicious of, advocated

by a person of whom almost all philosophers are suspicious. To make the underlying issues clearer, let me focus on another morally ugly character: Professor Philosophy, who employs his status to gain sexual "favors" from his female students, while self-satisfyingly and publicly insisting that the department conforms to an anti-harassment policy, and calling attention to his colleagues' violations.

I concur that Professor Philosophy is presumptuous for calling attention to his colleagues' shortcomings. Here, I think, we may well want to say that his repeated violations impact on his standing in the moral community;[21] as hypocrites are apt to do, he insists that others conform to rules that he ignores. He is free-riding; he is seeking to gain the benefits of living under a moral rule, but he is not prepared to bear the costs (compare Allessandra, version 1). Yet — and this is the crucial point — his readiness to call out his colleagues remains important, for life according to the rules is a public good, *as is bothering to enforce them.* A main claim of *OPR* is that modelings of stable cooperation indicate that a necessary precondition is the readiness of those not directly involved in a norm violation to help enforce it; if only the parties engaged in the transactions take an interest in the rules — if it is only their business, like Bob and his wife's coffee buying arrangements — cooperation among strangers is endangered in large groups. In my view that is why large-scale cooperation cannot be based primarily on a *modus vivendi*, an equilibrium of interests. If social morality was a sort of descriptive convention only held together by interests (or even a tit-for-tat like direct reciprocity) it would break down; the possibilities for successful defection are far too numerous. That is why self-sustaining moral constitutions give third-parties standing, for their interventions are moral public goods — all participants gain from their enforcement of the moral constitution. Indeed, there is strong reason to think that this idea of third-party standing goes

back to the very beginnings of the moral enterprise; in many ways it is its foundation.[22]

So the moral standing of Professor Philosophy looks quite complicated. He should be ashamed of, and guilty about, his violations, and if others are willing to effectively bear the enforcement burdens, then it perhaps would be seemly for him to wait until he has regained full standing in the moral community. On the other hand, if we wait to find someone without sin until the first stone is cast, we may find no stones are thrown, and violators flourish. What *do* we want Professor Philosophy to say at the departmental meeting? Do we wish him to remain silent, even if that affects the vote on the harassment policy? Do we wish him to abstain? If he sees Distinguished Professor Moral Philosophy harassing a women student, is he is being non-presumptuous, and so virtuous, if he passes by in silence? In any event, the important point is that *OPR*'s claim that all participants generally have standing to insist on your conformity is not a claim that all other people are as good as you (perhaps in many ways they are indeed morally lost, benighted, and blind), but that flouting the moral constitution is everyone's concern, and those who take the time to notice are fulfilling a crucial moral job. That when fulfilling this function they may be also be expressing vices (such as hypocrisy, or glee that another has slipped), should not blind us to the moral importance of recognizing that complying with the moral constitution is everyone's business.

6. Social Morality as a Collective Endeavor

Social morality is everyone's business because it is essentially a collective accomplishment that requires collective maintenance. Far too often, even in social and political philosophy, moral reflection and justification is assumed to be a highly individualistic inquiry and, yes, a difficult one and so inevitably elitist — far too

difficult for *hoi polloi*. Far too often social and political philosophy is envisaged as the expert philosopher perceiving the right; tablets inscribed with the truth (or pdfs from *Ethics*) are sent down from on high to the rest to guide them, but the philosopher demeans himself by descending and becoming a participant in the collective project that he cannot direct. And *hoi polloi* who do reason the matter through, and yet cannot grasp the truth of the message, are told that they do really do have reason to do as they are told, and this shows that the philosopher has addressed them as equals. *The Order of Public Reason* seeks to rediscover a different understanding of moral life. Its aim is not primarily to instruct, but to understand how in our diverse world social morality can be a collective achievement that responds to the normative commitments of the widest possible range of people, providing a structure that informs participants' intense social interactions in a cooperative society — a structure that is upheld by the free action of the participants themselves.

Notes

[1] John Rawls, "Kantian Constructivism in Moral Theory," in *John Rawls: Collected Papers*, ed. S. Freeman (Cambridge, MA: Harvard University Press, 1999): pp. 306-7. Emphasis added. The other essay is "The Independence of Moral Theory" in ibid., pp. 286-302.

[2] Rawls, "Kantian Constructivism in Moral Theory," p. 326.

[3] For a much more detailed account, see my essay "On the Appropriate Mode of Justifying a Public Moral Constitution," *The Harvard Review of Philosophy*, forthcoming.

[4] See P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy*, vol. 48 (1962): 187–211. For a view of social morality close to that of *OPR*, see his "Social Morality and Individual Ideal," *Philosophy*, vol. 36 (1961): 1–17.

[5] I explore this idea further in "The Egalitarian Species," *Social Philosophy & Policy*, forthcoming.

[6] See Paul Weithman, *Why Political Liberalism?* (Oxford: Oxford University Press, 2010), esp. chap. 2. On the distinction between stability and robustness, see my essay "The Role of Conservatism in Securing and Maintaining Just Moral Constitutions: Toward a Theory of Complex Normative Systems," *NOMOS: Conservatism* (New York: New York University Press, forthcoming).

[7] To be read: "is ordered above" or "is preferred to."

[8] This line of reasoning presents us a moralized version of Hobbes's fool. The fool wants the benefits of a social life according to settled rules, but whenever he can do even better by breaking them, he does so. See Hobbes, *Leviathan*, edited by Edwin Curley (Indianapolis: Hackett, 1994), p. 90 (chap. 15, ¶4). Allessandra seems a moral fool.

[9] Immanuel Kant, *The Metaphysical Elements of Justice*, 2nd edition, edited and translated by John Ladd (Indianapolis: Hackett, 1999), p. 116 [§43]. Emphasis added.

[10] On some views living according to the moral constitution will never depart from personal moral optimization: e.g., a moral view according to which what is morally best is for you to follow the society's eligible morality. A rule consequentialism might be designed in a way to do this (compare a Hobbesian whose greatest desire is to obey the sovereign). That some views can avoid the conflict does not show that the conflict is not basic to social life.

[11] John Rawls, *Political Liberalism*, expanded edn (New York: Columbia University Press, 2005), p. 51.

[12] For skepticism of suppositions about rational moral monsters, see my paper "On Theorizing About Social Morality*," European Journal of Analytic Philosophy*, forthcoming.

[13] T. H. Green, *Lectures on the Principles of Political Obligation*, in *Lectures on the Principles of Political Obligation and Other Writings,* edited by Paul Harris and John Morrow (Cambridge: Cambridge University press, 1986), §105.

[14] I think that Talisse's opponents of stem-cell research are making much the same point. I regret that space limitations do not allow consideration of important worries from the perspective of religious citizens raised by Talisse. Happily, Kevin Vallier's new book carefully and comprehensively analyzes these issues — and with much more knowledge than I possesses. See his *Beyond Separation: Uniting Liberal Politics and Public Faith*, (New York: Routledge, 2014).

[15] Hobbes, *Leviathan*, p. 25 [chap. V, ¶3].

[16] For a fascinating study of the place of the authority relation in our attribution of moral responsibility, and what may be left of our normative relations with those who cannot make sense of it, see David W. Shoemaker, "Psychopathy, Responsibility, and the Moral/Conventional Distinction," *The Southern Journal of Philosophy* 49, Special Supplemental (2011): 99-124.

[17] I am a bit perplexed that Quong thinks this result is intimately related to my objection to the "bracketing strategy." My complaint about the bracketing strategy is as a *strategy for justification* of a moral constitution. If we say that rule *R* is justified to Alf just because it satisfies a subset of his relevant reasons, then (*i*) we have only offered a *pro tanto* justification, since relevant reasons were excluded and (*ii*) if we

insist that the *pro tanto* justification is sufficient, we do not have strong grounds for thinking that if *R* is justified to Alf he will tend to comply with it, for in our "justification" we have bracketed many of the reasons, even morally relevant ones, that matter in his decision to comply. This is a problem of *inherent* stability. In contrast, the result concerning Carl does not concern someone to whom *R* is said to be justified, but people living among us *to whom we acknowledge it is not.* This raises a question of *external* stability: how can *we* stabilize a scheme when some with whom we interact reject it? On some of the problems of relating public justification and stability, see Paul Weithman, "Relational Equality and Inherent Stability," *Social Philosophy & Policy*, forthcoming.

[18] Adam Smith, *The Theory of Moral Sentiments* (Indianapolis: Liberty Fund Inc., 1982 [1790], pp. 176-7. Chad Van Schoelandt has taught me a great deal about this issue; it was he who alerted me to the importance of this passage of Smith's. Van Schoelandt has analyzed the problems raised by Carl in much more depth, and with much more subtlety, than I have been able to here. See his insightful "Justification, Coercion, and the Place of Public Reason," Working Paper, University of Arizona Philosophy Department.

[19] I am grateful to Paul Weithman for getting me to think more deeply about this issue.

[20] Compare the unsympathetic Carl to President America, who holds that the norm against killing civilians without trial has no authority over him when he decides other values outweigh it.

[21] Geoff Sayre McCord has explored this thought — that violations may cause one's standing in the moral community to temporarily lapse. See, for example, his "Criminal Justice and Legal Reparations as an Alternative to Punishment" in *Legal and Political Philosophy*, edited by Enrique Villanveua (Amsterdam: Rodopi, 2002): 307-38.

[22] See further Gaus, "The Egalitarian Species."