ON THEORIZING ABOUT PUBLIC REASON

GERALD GAUS

University of Arizona

ABSTRACT

This essay responds to the thoughtful essays on the *Order of Public Reason* (*OPR*) by Elvio Baccarini, Giulia Bistagnino and Nenad Miscevic. All three essays interrogate *OPR*'s understanding of moral theory - "meta" matters about the nature of morality, reasons and modeling within moral theories. I first turn to the general understanding of the moral enterprise underlying *OPR,* explaining why it takes a view at odds with the contemporary mainstream in moral philosophy. I then explain the idea of moral truth in *OPR:* when it comes to social morality, moral truth is necessarily a function of what can be endorsed by some collectivity. This leads to a fundamental worry about theories of public reason: why is the endorsement of the public so important? And if some sort of public endorsement is really so terribly important, how can a theory of public reason withstand the fact that it advances its own controversial claims that cannot be publicly endorsed? After considering when public endorsement is necessary, and when public reason theories can make controversial claims, I close by considering in what way *OPR* does, and in what way it does not, employ a thought experiment, and the complexities of that experiment.

Keywords: public reason, social morality, reactive attitudes, mental experiments

1. The Layers of Public Reason Moral Theory

I am most flattered by the attention and care given to *The Order of Public Reason* (*OPR*) (Gaus 2011) by Elvio Baccarini, Giulia Bistagnino and Nenad Miscevic. Although we disagree on many matters, throughout their essays they raise important questions. I cannot even attempt to answer all their queries or respond to all their worries. In any event, I am not confident that it would be a good use of the journal's pages, or the reader's time, to attempt point-by-point rebuttals. I have myself read an essay or two that seeks to engage in point-by-point critiques of *OPR*, and the result strikes me as unsatisfactory. There is bound to be misinterpretation, over-defensiveness (or aggressiveness), and possible replies and solutions left unexplored. And it all gets rather tedious rather quickly. What we should seek is a deeper understanding of important issues, not to score points in a game. So I propose to explore some of the central issues raised in Baccarini's, Bistagnino's and Miscevic's thoughtful essays, sketching their concerns and explaining how I approach these matters. Although at some points I mark where, in my view, they have gone wrong, I make no claim to rebut or refute their views; any fair attempt to do so would take a very long conversation, and even at the end much would be left unresolved.

I note that Bistagnino gently chides me for being harsh with alternative conceptions of moral and political philosophy. And it is indeed true that I believe that much contemporary moral and social philosophy is ill-conceived; *OPR* explicitly calls for a reconsideration of the aims and scope of moral and political theorizing. So it is helpful that all three essays interrogate *OPR*'s understanding of moral theory — "meta" matters about the nature of morality, reasons and modeling within moral theories. I shall follow their lead and restrict attention to my conception of moral theorizing — how I conceive of social morality and how we should theorize about it. This means that I shall set aside interesting and important substantive issues about egalitarian liberalism and socialism. I do regret that.

We are well acquainted with the distinction between normative ethics and metaethics, but that distinction does not exhaust the different layers of a moral theorizing, especially in a public reason theory. Consider an incomplete list. First, a public reason theory such as that presented in *OPR* must identify a set of rules, institutions or practices that are the subject of analysis. In *OPR* this concerns social morality and political institutions. This is an issue about the *subject* of moral inquiry. I turn to this in section 2, where I try to explain why *OPR* takes a view of these matters at odds with the contemporary mainstream in moral theory. We next might ask what we hope to achieve from moral inquiry — the *goal* of our theorizing. This may seem obvious to many moral philosophers: the aim is to settle on the truth about morality. But like most simple formulas, this one belies fundamental differences. Section 3 seeks to explain the idea of moral truth in *OPR*. Traditional philosophy has often sought truth through intuitive reflection, according to which a person can come to justified true belief about morality through reflecting on her deepest convictions, abstract principles or, perhaps, constructing mental experiments that test principles in some set of conceivable cases. Alternatively, moral truth is sometimes understood as constructed out of one's personal moral convictions. In any event, *OPR* holds that when it comes to social morality, moral truth is necessarily a function of what can be endorsed by some collectivity. This leads to a fundamental worry about theories of public reason, explored in section 4: why is the endorsement of the public so important? Even if it is *a* consideration, why does *OPR* treat it as definitive? And if some sort of public endorsement is really so terribly important, how can a theory of public reason withstand the fact that it advances its own controversial claims that cannot be publicly endorsed? Section 5 explains in what way a theory of public reason can make controversial claims. Lastly, like Rawls' magnificent *A Theory of Justice*, *OPR* employs a device to translate the abstract idea of public justification into more concrete conclusions; I call this the Deliberative Model, the subject of section 6. Responding to the Miscevic's fascinating essay, I consider in what way *OPR* does, and in what way it does not, employ a thought experiment, and the complexities of that experiment.

As if all of this did not sufficiently multiply the number of distinct issues and levels of analyses, there is another, deeper, question: what is my view of the moral enterprise that motivates the work? Some answers to this question are part of the internal commitments of *OPR*, while others concern my own, much more controversial, pretty thoroughly naturalized, view of the moral enterprise. The latter concerns are, I trust,

consistent with *OPR*, but not presupposed by it, since I stress that those who do not share many of my deep commitments about the moral enterprise can, and should, take the view of public reason about social morality. In *OPR* I tried my best not to bring these in; here I allow myself the luxury of some appeal to them. This may help the reader to understand me; I hope it does not confuse her about the book.

2. Social Morality: The Distinctive Collective Human Achievement

The subject of *OPR* is the justification social morality: unless we can at least get clear about the idea of social morality, we will be talking at cross-purposes. Many contemporary social philosophers, especially those who come to the field via metaethics or abstract normative ethics, do not even see social morality, though I believe that they are embedded in it, and it informs their lives. It would not be a terrible exaggeration to say that it has became largely invisible in late twentieth/early twenty-first moral philosophy, and much of *OPR* is devoted to trying to get us to see it again.[1] I say "again" because earlier moral philosophers such as Hume, Hegel and Mill, and more recently Kurt Baier (1995) and P.F. Strawson (1961), saw it clearly. And so did John Rawls.[2] But though Rawls focused on it, most of his interpreters and students did not; it remained, if not totally invisible, shrouded to them. Today, many philosophers have adopted a view of the moral enterprise according to which social morality simply cannot exist. It is no wonder why some seem so perplexed by *OPR*.

In her contribution Bistagnino is clearly puzzled by the idea of social morality. Drawing on David Enoch, she writes:

> This is the riddle I wanted to highlight: in OPR Gaus wants to propose an innovative and original understanding of morality by defending a combined approach between Humean and Kantian understanding of it; within such a view, rules of social morality are somehow both positive and moral. However understanding such a precarious equilibrium between the descriptive and the normative is not easy *for it seems that the rules of social morality cannot be both normative and positive.* If the rules of social morality are normative because they pass the test of the moral point of view, *it is irrelevant whether they are embedded in a society for such an embracement is only contingent.* On the other hand, if rules of social morality are those that are actually followed in society, it is not clear in what sense they are moral. The only way out of this riddle would be to endorse some sort of philosophy of history apt to show that human progress corresponds to moral progress [emphasis added]. (Bistagnino 2013, 19-20)

I hope that in the next two sections it will become manifest why I do not see any riddle or even an equilibrium, much less a precarious one (and so why I certainly do not endorse a philosophy of history).[3]

---

1 I (2013*a*) make this point at greater length in "On Being Inside Social Morality and Seeing It."
2 I (forthcoming*a*) have argued for this interpretation in "On the Appropriate Mode of Justifying a Public Moral Constitution."
3 Recently I have tried to analyze the relation between social evolution and moral improvement in a way that avoids appeal to the idea that somehow social evolution is morally progressive. The matter turns out to be rather complicated. See my 2013*c*. "The Evolution, Evaluation, and Reform of Social Morality: A Hayekean Analysis."

Let me begin by stepping back to propose an analysis of the heart of the problem — why the idea of social morality seems to be a riddle to so many contemporary philosophers.[4] I suspect that Bistagnino and many others have three basic theoretical convictions about the moral enterprise that *OPR* rejects.[5] Perhaps the most fundamental is:

> *The Strict Positive/Normative Distinction:* Whether or not Alf truly morally ought (or ought not) to φ (or whether he truly has a moral right *R,* or a moral duty *D*) in society *S* does not (in any important way) depend on whether there is an actually recognized social rule in *S* according to which Alf ought to φ (or has a moral right *R,* or a moral duty *D*).

As stated, this is rather vague (and I don't think I can do much better). I add "in any important way" as even the most orthodox of moral philosophers might allow that moral duties sometimes derivatively refer to conventions (say, a moral duty to drive on the right in continental Europe, but on the left in the UK). But, overwhelmingly, moral and political philosophers today are convinced that moral propositions, moral imperatives, rights and duties are essentially independent of actual social rules, as the Strict Positive/Normative Distinction requires. Conventions and social practices, of course, should conform to the demands of true morality, but the mere fact that the duty to φ is not embedded in an actual social rule cannot (except in the sort of case indicated above) undermine the truth of the claim that it is morally required that one φs.

The second basic conviction rejected by *OPR* is that, in a deep sense, conclusions about true morality are matters of individual judgment. We can call this:

> *The Strong Moral Autonomy Conviction*: Each competent moral agent in society *S* properly arrives at her own judgment as to what morality truly requires.
>
> *(i)* The justification of this judgment does not require reference to any collective determination, decision, or social fact as to what the morality of society *S* is, nor
>
> *(ii)* does the correct answer to what morality truly requires depend on what the person inquiring thinks it requires, or what the other people in *S* think it requires.

The Strong Moral Autonomy Conviction does not require that each autonomous agent make up her mind about morality in isolation; she certainly may consult, and deliberate with, others. But if she does so these are merely inputs into her own decision process, for ultimately she must arrive at her own judgment of what morality requires, and this judgment does not necessarily refer to any collective determination. Clause (*i*) is important since *all* beliefs of a person ultimately depend on her own judgment.

---

4   Some have even suggested that I invented (or perhaps they would say, concocted) the idea. I am honored, but, alas, Hume beat me to it; I learned it from the works of one of my teachers, Kurt Baier.
5   Much of this section draws on my (2013*b*) essay, "Why the Conventionalist Needs the Social Contract (and Vice Versa.)" That essay was written partly in response to this concern of Bistagnino's; I cite her essay there.

Even if social morality were, say, the outcome of a vote, a person's judgment about it would depend on *her* beliefs about the outcome of the vote. Ultimately, there is no way to escape the point that one's judgments are based on one's beliefs. The Strong Moral Autonomy Conviction insists that not only (of course) is this so, but in coming to this judgment it is not necessary to include any beliefs or suppositions about what we, collectively, have arrived at as our morality.

Clause (*ii*) is intended to capture the idea that a morally autonomous agent's inquiry into morality aims at coming to a conclusion about a matter the correctness of which is independent of this inquiry and her beliefs about it. She does not think her believing that "morality requires φ," or that the flawless use of her decision procedure has led to the conclusion that "morality requires φ," *constitutes* the truth of the claim. The Strong Moral Autonomy Conviction insists that morality is an object of inquiry, which is not itself constituted by anyone's inquiry or decision procedure. What morality requires does not depend on what I think it requires although, of course, *my conclusions* about what it requires do most definitely depend on how I see it. Again, there is no possibility of getting outside of one's beliefs.

The third and last conviction at the heart of most orthodox moral theorizing is:

> *The Anti-technology Conviction:* Morality is a not a technology to enable human cooperative social life. Morality does not have necessary functions:
>
> > (*i*) it simply *is*, though it also
> >
> > (*ii*) instructs us what to do.

OPR is based on a rejection of these three philosophic convictions. Consider the last. As Bistagnino properly recognizes, *OPR* (in common with philosophers such as Philip Kitcher (2011, esp. chaps. 2 and 6)) understands social morality as a technology for human cooperation that is, perhaps, *the* innovation that made humans the eusocial creatures we are. In writing *OPR* I sought to bring at least a critical part of morality down from the heavens — agreeing with David Gauthier (1991) that a defense of morality cannot *presuppose* belief in an enchanted universe, populated by invisible properties that instruct us what to do. Instead, *OPR* sees morality as the distinctive and perhaps crowning innovation of our species, which enables us to be the types of creatures we are. Like Darwin (2004 [1879], 120), I "fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important." Morality is the supreme human adaptation. For a surprising number of philosophers (indeed, I fear most), even if no one ever believed, or acted upon, the conviction that we have a moral duty to φ — indeed apparently even if humans were a very different sort of species so that *no one would ever* φ — it could nonetheless be true that we have a moral duty to φ.[6]

---

6 G. A. Cohen (2008, 20) insists that the infeasibility of a vision of justice does not "defeat the claim of a principle." David Estlund (2011) also defends the relative independence of the demands of justice from our natures.

Those who embrace the Anti-Technology Conviction deny that morality is, at bottom, a human innovation that was, and continues to be, a solution to a fundamental problem of social living among primates like us. To a philosopher of this ilk, morality just *is*, and it commands us what to do (through her, of course). This last point is important: one who has the Anti-technology Conviction can, and typically does, maintain that social morality entails prescriptions and a notion of the optimal rules of social organization. Bistagnino is entirely right that one who follows G. A. Cohen holds that moral truth has implications for practice and social organization. But this is very different from seeing morality as itself a tool or device that humans have hit upon, for on this latter view what is the *true* morality (see section 3 below) intimately concerns what moral code can suitably fulfill the necessary functions. It is not that the truth entails implications for practice, but that what is true is that which can serve practical functions in suitable ways.

Because so many philosophers have this anti-functionalist understanding of the moral enterprise (so they do not see it as an inherently *social tool*) they are attracted to a highly individualistic conception of moral judgment, as stated in the Strong Moral Autonomy Conviction. Even if morality is about what *we* must do (even if it says that *we* must live under socialism), *I* judge this; it is no way a collective judgment. It is this radically individualistic conception of social morality that the social contract tradition (from which public reason views derive) dismissed.[7] Even if in some contexts each moral agent has a definitive say about the normative requirements that apply to her, this is not the case when it comes to our shared social existence, which is also a *shared moral* existence. Recall that even Kant described the state of nature as one characterized by violence because each claims "the right to do what *seems just and good to him*, entirely independently of the opinion of others." (1999, 116 [§43], emphasis added)

We each have views about what is just, but social life is a collective achievement; the crux of the social contract tradition is that we need a collective determination of the nature of a just social life. To be sure, at the same time, we each have our own ideas about what is just and moral, and an autonomous person cannot forsake the use of her own judgment. *OPR* thus can be seen as resting on:

> *OPR's Moral Autonomy Commitment:* A reflective moral agent in society *S* properly arrives at her own judgment as to whether the moral rules of her society warrant her internalization of them, and the adoption of the other moralistic attitudes (guilt, resentment, etc.).
>
> > (*i*) The moral rules must sufficiently cohere with her personal normative convictions, or at least not conflict too much with them.
>
> > (*ii*) Rules that pass this test are endorsable.

OPR's Moral Autonomy Commitment is not so extravagant as the stronger version we examined. Autonomy does not require that each come to her own definitive conclusion

---

7   I argue this in "Public Reason Liberalism" (forthcoming*b*). For a much more comprehensive picture, see *The History of Public Reason in Political Philosophy* (forthcoming*c*).

about what morality requires independently of the judgment of others, but that an autonomous agent reflects from her perspective on the endorsability of the rule that we collectively have arrived at. One way we might understand the fundamental difference between the two formulations of moral autonomy is to distinguish optimality from acceptability. The Strong Moral Autonomy Conviction identifies moral judgment with a sort of optimality: one justifiably believes that "morality requires φ" if one believes it is the best answer to the moral problem, the one that is most coherent, or the one that based on the principles that are in in "reflective equilibrium" for you. The more modest conviction on which *OPR* rests does not suppose that *any* personal judgment that φ is the optimal answer justifies belief that morality requires φ, for to be required φ must also be part of a socially-maintained convention. But more than that, as clause (*i*) of *OPR*'s Moral Autonomy Commitment stresses, a morally autonomous person does not require that her society's moral norms perfectly correspond to her own individual moral reflection about optimality, but only that they are not opposed, or not too deeply opposed, to her normative perspective. Sharing a cooperative social existence under moral rules is a great human good; a morally mature person does not reject the enterprise because it does not perfectly correspond to her own controversial judgment about optimality.

We are now, I think, in a better position to appreciate

> *OPR's Positive/Normative Distinction*: That according to the positive conventional morality of society *S* Alf ought (or ought not) to φ (or has a moral right *R* or duty *D*) does not imply that Alf really morally ought (or ought not) to φ (or has a *bona fide* moral right *R* or duty *D*) in *S*.

Social morality can only fulfill its function if its rules are actual social rules; even the most ideal non-actualized rules cannot form the groundwork of social existence. But this does not imply that moral rules are *only* actual social rules. They must be actual social rules that are consistent with the moral autonomy of its participants — which are endorsable by them. The ability to stand back from our social rules, and decide whether they correspond to one's idea of an acceptable way of living together, is fundamental to being an autonomous person. Those indoctrinated into social arrangements such that they cannot question them, but simply internalize and enforce them just because that is the done thing in their group, are indeed heteronomous: they are dominated by the rules of their group.

3. True and Absolute Moralities

Following the general usage suggested by Baier, we can state, as an approximation

> *OPR's Truth Claim*: In society *S* at time *t*, it is true that "There is a moral duty to φ in circumstances C" if and only if
>
>> (*i*) in *S* at *t* there is an actual social rule *M* according to which "There is a moral duty to φ in circumstances C";
>>
>> (*ii*) *M* has the features of moral rules and

(*iii*) *M* is endorsable by all members of *S* at *t*.

This not an arbitrary specification. In this case *M* has the formal characteristics of a moral rule (e.g., generality, publicity) and so can be distinguished from other social rules. Because it is an actual social rule it can actually provide the shared basis of expectations and social order, and so serves the function of morality. And it is endorsable by all, and so expresses the moral autonomy of each. There is no equilibrium or precarious compromise; it is closer to the relation of a set (social rules that have the characteristics of moral rules) and its proper subset (those that are endorsable). On the Strict Positive/ Normative Distinction this does not make sense: the "ought" or the normative cannot be constrained by what "is," the empirical. But that *given* certain presuppositions a view like *OPR*'s does not make sense only condemns the view if those presuppositions are unassailable. But as I have indicated, not only do I not see them as unassailable, I believe them to be deeply mistaken, reflecting an overly individualized view of the ethical project that does not appreciate how it was the creative adaption that was critical in transforming a very intelligent, but still pretty standard primate, into those hyper-social creatures we call "humans."

Once we set aside the three orthodox convictions, then the "moral" is not in tension with the "empirical," though it is certainly not reduced to it. Moral truth is not some sort of external target that the social rules of a society aim at; if that were so, then Bistagnino might be correct that I would need a philosophy of history to explain how these actual rules come to hit their target. But the "truth maker" for claims in social morality, based on an actual rule *M,* is the endorsability of *M* by the members of *S*, given whatever normative commitments and beliefs they happen to have at *t* in *S*. Thus, as Baier stressed, there can be more than one "true morality" (or different moral truths in different societies), through there are also many false moral claims in all societies.

Now, of course, we can ask of some proposed rule *M** (a rule that is not yet a social rule): *if M** *were* a social rule, would it be endorsable? And if we answer in the affirmative, this is an important moral conclusion, but it would not make *M** into a moral rule, as *M** cannot perform the functions of a moral rule. We might think of *M** as a blueprint for a moral rule; no matter how wonderful a blueprint is, it does not do the work of the "built" technology. We can say that it is true that it is a good blueprint, but it does not yet provide a built technology that forms the basis of normative expectations; in the strict sense it does not generate rights, duties, and obligations.

There are numerous reasons for not treating rule blueprints as if they were rules; they are not generally known, there is no empirical expectation that others will act on them, and of course we never quite know what will happen when a blueprint is translated into a technology. One reason for this is, as Hayek (1973, esp. chaps. 2-4) points out, a linguistic statement of a rule is always approximate, and leaves out many aspects implicit in its actual operation. Rather than "blueprints" we should, perhaps, talk of schematics. But more importantly, there will almost always be alternative, inconsistent, proposed rules that are endorsable (and we have no uncontroversial way to socially rank them); if each rule that was endorsable *ipso facto* became a *bona fide* moral rule, we would end up with numerous conflicting moral rules. Thus, *OPR* stresses that the most

we can say about our schematics is that in society *S* there is an endorsable prescription that there be a moral rule about such matter in *S*, and that the members of *S* ought to work to implement one if they do not already have one. This is what I called a *moral principle.* Yes, it is normative, but it does not operate as a technology of cooperation; it provides an schematic for a variety of related rules, and tells us to work on creating an actual rule that meets its general specifications.

Now suppose that there is some principle that we think is endorsable in all human societies, or at least any human society that is relevant for our moral thinking today. We can call these *human rights* (Gaus 2011, 180, 428ff).[8] As I stress in *OPR*, a human right is not a *bona fide* moral right, but more of a call to work for the institution of a type of right, along with our conviction that a society that does not have any right on this matter is morally defective (because it fails to satisfy a justified moral principle). We can call a society that does not honor "human rights" morally defective, oppressive and even "unjust." But we should not confuse ourselves into thinking that such schematic rights can perform the function of the built technology — moral rights. One of the things we have learned is that the real value of human rights discourse is not in the philosopher's claim that there *is* such-and-such a right and his mental experiments to specify it, but in the implementation of international norms and social rules that can actually provide the coordination of expectations and actions.

OPR calls these principles of human rights *transcendental moral claims*: claims that every society must take these vague schematics and develop some moral technology that meets them. And that is because we believe they are endorsable in every society.[9] Baier called them "true" but that seems misleading, since the crux of the doctrine is that we do not really have *bona fide* moral rights and obligations until they are actualized in rules. *OPR* uses the term "transcendent" to indicate that their endorsability is not restricted to specific societies.

So I must confess that, *pace* the much-cited Enoch (quoted in Baccarini, note 1), I do not see this as giving any game away.[10] Quite the contrary: it is a straightforward implication of the possibility of endorsing schematics conjoined with the idea that the endorsability of some schematics is part of all human social orders. And, of course, I hope it is clear that, despite my use of the terms "transcendent" and "absolute," there is nothing very Hegelian about the idea.

4. Endorsability and the Reactive Attitudes: The Intrinsic Good of Public Justification

This whole account turns on "endorsability." Public endorsability — or public justification — is necessary to social normativity. Now what is so special about public

---

8   I also consider the idea of human rights in my 2014 contribution to a symposium on *OPR*.
9   Or, more modestly, that we are sufficiently skeptical of claims that in some societies they are not endorsable so that we act on the supposition of universal endorsability. Or, at a *very* minimum, we so empathize with those free and equal persons who do endorse them than we give a strong presumption to the claim that, if the others seriously considered these people, they would endorse them (Gaus 2011, 430).
10 Although I employ game theory, I do hope that, in the end, philosophical investigation is itself more than a game. But one sometimes wonders.

justification or endorsability: why do public reason accounts place it at the very heart of social morality? This immediately leads to one of the issues Baccarini discusses: the place of the reactive attitudes in my account. Let me first explain the important role of the reactive attitudes in *OPR*, and then turn to some of Baccarini's concerns.

We can imagine two different types of public reason theories that rest on two different answers to the foundational question: why engage in public justification?[11] We might call the first an act–constraint account. On this view there is some sort of act φ, which is permissible only if φ can be publicly justified, where φ is an act that can occur without justification. That is, public justification is not needed to φ, but to permissibly or non-wrongfully φ. So here public justification seems to be a straightforward normative standard according to which one may not permissibly φ without public justification. Such accounts immediately raise the question of the status of this normative standard: is it a "true" normative standard that somehow reflects a deep realist moral commitment (and so relies on the Strict Positive/Normative Distinction after all)? It also raises what we might think as *the public justification fetish* worry: surely if φ is something *very* important, appealing to a normative standard that requires public justification for any permissible φ-ing looks like placing overwhelming weight on a relatively unimportant issue: is φ endorsable by all? Perhaps it would be a good thing to show that everyone endorses φ, but can it really be a necessary condition for permissible φ-ing?[12]

I certainly am not saying that these problems cannot be solved, but they are not foundational problems confronting *OPR*. A distinctive feature of *OPR* is that public justification is intrinsically necessary to achieve certain goods — the goods of maintaining the reactive attitudes of resentment and indignation, guilt, and our practices of moral responsibility. We cannot (*not* "may not") achieve these goods without justification (as we can coerce without justification). Absent endorsability the good simply cannot be achieved. The key here is:

> *OPR's Strawsonian Claim:*
>
>> (*i*) The reactive attitudes and allied practices of moral responsibility reflect the attitudes of participants in a practice;
>>
>> (*ii*) when Alf experiences these attitudes towards Betty for her φ-ing, he is reacting to his belief about the quality of her will or the characteristic of her perspective such that
>>
>> (*iii*) it reflects an ill-will or ill-regard towards Alf.
>>
>> (*iv*) But this implies that Alf in criticizing Betty supposes she must have been able to appreciate that she ought not to have φ-ed, but did so anyway; otherwise her failure to refrain from φ-ing is not a form of ill-will to, or contempt for, Alf;

---

11 In thinking about these matters, I have greatly benefitted from discussions with Chad Van Schoelandt.
12 Van Schoelandt forcefully argues that if φ = coercion, this question cannot be plausibly answered in the affirmative. See his "Justification, Coercion, and the Place of Public Reason," Working Paper, University of Arizona Philosophy Department.

(*v*) and so we can say Alf must suppose that Betty can reason herself to the conclusion that she ought to not have φ-ed.[13]

If the supposition in (*v*) is not met, Alf cannot claim that Betty's φ-ing manifests ill-will or ill-regard towards him. Perhaps she is a newly arrived foreigner who does not know of the rule about φ-ing; or perhaps she is a thoughtful and conscientious person who simply cannot see any good reason not to φ, and sees compelling reasons to φ. According to *OPR*'s Strawsonian Claim, the reactive attitudes are the result of looking inside the other — reading her mind — and determining whether the rule we invoke is one she can see as something that she should conform to; otherwise we cannot react to her ill-will, for then there is no ill-will.

Given this Strawsonian analysis, to possess a social morality in *S* that rationally sustains the reactive attitudes and our practice of moral responsibility (i.e., in a way that does not depend on having unjustified beliefs), it must be the case that, when rationally deliberating about the violations of others, we believe that they did, or at least could be reasonable expected to, see the point of these rules and endorse them (again in a way that does not require them to make errors in reasoning, or be hoodwinked into manifestly unjustified beliefs). If that is so, their violations evince Strawsonian ill-regard, and the reactive attitudes can get a grip. But this is essentially a requirement that the rules be endorsable by all those in *S* subject to them. And so we are led to public justification as necessary in securing the fundamental goods of sustaining reactive attitudes and our practices of moral responsibility.[14]

At this point Baccarini dissents:

> The account that Gaus offers does not correspond to a reflection on our moral practice, but to a controversial philosophical position, if, among other, it includes an ambiguous or contested conception of having a reason, because this conception plays a crucial role in the account. Another possible reason to worry about Gaus' intention to offer a reflection on our moral practice, as opposed to a controversial philosophical position, appears if the reflection does not correspond to our moral phenomenology…. (Baccarini 2013, 27)

I confess to being not entirely clear about what constitutes an inquiry into moral phenomenology; I take that it involves a sort of introspection as to how an experience feels, or looks, from the inside. While an introspective method can reveal the immediate quality of an experience, it is far less useful in getting at the underlying causes and presuppositions. One does not have to be a Freudian to suppose that we often are not aware, or at least not fully aware, of the beliefs and presuppositions that an experiential state rests on. In these cases we must appeal to other sorts of data as well as theories of the phenomenon. The Strawsonian account of *OPR* rests on two main sources.

First, I believe that the body of recent philosophical work on moral responsibility

---

13   See Strawson (1962), "Freedom and Resentment."
14   This is not to say that these are the only goods, but a system of morality that did not almost always sustain them would be alien — it would be a morality without responsibility.

endorses some version of the Strawsonian analysis. To be sure, there are subtle differences in the details of the accounts, and the type of approach I favor is one of several in play. But these are relatively minor matters about more precise details; the important point is that Strawsonian analyses are central to current discussions; they justify focusing on the internal, participant, perspective.

Secondly, my early *Value and Justification* (1990) surveyed theories of emotions and attitudes and relevant data. I believe it quite clearly showed that the most compelling accounts of emotions view them as cognitive-affective complexes (I believe this thesis has become even better supported as the years have passed). An emotional state has both an affective felt component (which is what is clear from introspection) and cognitive elements. Without the appropriate cognitions, an emotional state cannot be maintained. Thus one simply cannot be afraid of α under the description "It is cute, not at all dangerous, and fun to play with." If that is what one really thinks, the fear of α cannot exit; if one changes one's beliefs about an object, affects adjust. To be sure, a person *might* insist that his introspection shows that the felt experience regarding α really was fear, even though he thinks its target is simply cute. That, he might say, simply *is* his phenomenology. However, we cannot let things end there. We need to inquire what is going on, despite the report of the phenomenology. Perhaps, as can happen, the person has misdescribed his emotion, confusing, say, being *surprised* at the appearance of a kitten suddenly popping out from under a blanket with being *afraid* of it. Or, if the person insists that he is really afraid, we may look for beliefs of which he is not aware, as Freud so often did. The crucial point is that a moral emotion is an affective-cognitive complex. If we combine this view of the emotions with the Strawsonian analysis, we are led to the conclusion that moral emotions such as resentment not only depend on the belief that the person did wrong, but the conviction that this state is an indication of ill-will or ill-regard. To be sure, this is a theoretical claim, supported by evidence and analysis, and as with every interesting theoretical claim it can be interrogated and questioned. But it cannot be refuted by mere introspection, reporting the felt experience and perhaps some of one's beliefs.

In any event, on to Baccarini's more specific criticisms:

> It seems implausible to take Gaus' internalism as a credible account of our moral practice if it importantly diverges from our moral phenomenology. Certainly, there are relevant counterexamples to strong externalism coming from moral phenomenology. It is implausible to blame a person who lived in Athens in the fifth century BC for accepting slave ownership. But there appear to be counterexamples to Gaus' account, as well. Certainly we blame and feel indignation for Goebbels, although it does not appear how our moral reasons could possibly enter in his evaluative standards. We blame and feel indignation for his wife for having killed their children, although this may perfectly follow from her evaluative standards. I am tempted to say that we blame religious parents who withhold medical treatment and let their child die because she did

not receive medical support.[15] (Baccarini 2013, 32)

I seldom have interchanges about my work without having to discuss Nazis, and so let us consider Goebbels. One radical way of reading Baccarini's remarks about Goebbels being a counterexample based on "externalism" is to attribute to Baccarini:

> View I. (*i*) It is a fact of the matter that Goebbels violated a moral rule;
>
> (*ii*) The belief in (*i*) is sufficient to ground a coherent emotion of resentment, and to sensibly hold Goebbels responsible.

I do not think that View I can be correct. We would certainly have to add considerations about the absence of various excusing conditions. I think it is now widely accepted that if Goebbels was a psychopath he would not have been morally responsible, for he could not understand the nature of the wrongness of what he did. In important ways, psychopaths just do not get the idea of morality. But notice now how we must, as Strawson says we must, depart from the external, objective, perspective and see how things look to the agent subject to evaluation. We must know things about *his view*, not just about external facts of the matter from our point of view. In addition to standard excusing conditions, I think that we need to hold something along the lines of:

> (✳) Goebbels either (*a*) believed he was doing wrong, or (*b*) could have come to that conclusion.

If something like (✳) does not hold, Goebbels would still have been a disaster for mankind — a moral moron or ignoramus who spurred countless horrors, but without knowing they were moral horrors. Perhaps he knew that people called him immoral, but he simply could not reach that conclusion. That, though, would not be enough to hold him morally responsible, and so to ground the reactive attitudes. To lock him up, yes.[16] For moral responsibility we need to have some beliefs about how things looked *to* Goebbels, we need to think about his participant perspective. Notice that something along the lines of (✳) states an *access* condition: there must be some sense in which the target of the emotion could have, but failed to, reason himself to seeing what is morally required of him. Thus, as Baccarini accepts, we do not hold the ancient Greeks responsible for slave owning if we think the idea that it was morally wrong was something the good use of their reason could not get them to see. But I do not see why with Goebbels we are not also making claims about access, and I see considerable arguments that we must.

View I thus looks pretty implausible to me. An argument that accepts the access condition but still seeks to reject *OPR*'s account would be:

> View II. According to *OPR*'s analysis:
>
> > (*i*) To sustain the reactive attitudes towards Goebbels for his participation

---

15 There is a dispute here concerning what are reasons and "internalism," which I defer until the next section.
16 David Shoemaker (2011) distinguishes the conditions for moral and legal responsibility in his excellent essay, "Psychopathy, Responsibility, and the Moral/Conventional Distinction."

in slaughter one must believe that Goebbels' actions expressed an ill-will that is relevant to moral appraisal.

(*ii*) One can hold that Goebbels role in the slaughter constitutes an ill-will relevant to moral appraisal only if one thinks:

> (*a*) he believed it was something he ought not to have done, o*r*:

> (*b*) he could have reasoned himself to that conclusion by level of good reasoning *l*, and our social expectations are such that we expect him to reason to level *l*. In this case too, Goebbels' participation in the horrors would have been the result of ill-will (we might say, he did not exercise due diligence);

(*iii*) Goebbels did *not* believe that he did wrong though he *did* engage in level of good reasoning *l* and rationally concluded that it was permissible to slaughter innocents.

But:

(*iv*) One nevertheless may sustain the reactive attitudes towards Goebbels, so there is a contradiction in (*i*)-(*iii*).

There are two ways that I would resist this conclusion. First, I would have to know a lot more before I became convinced that Goebbels exercised the required level of good reasoning and yet reached the conclusion that he did no wrong. To accept II.iv we need to accept II.iii: we need to assure ourselves that Goebbels' system of values did not really imply (at level *l* of reasoning) that he was doing wrong. I am not at all confident that his system was so insulated from reasonable moral conclusions. Consider:

> We do not want to abolish pleasure, but rather to let as many as possible share in it. That is why we encourage people to attend the theater, that is why we give workers the opportunity to dress well for festive occasions. That is the reason behind *Kraft durch Freude.* That is why we shake off the agents of a prudish hypocrisy, why we do not allow decent, hard-working people who have every reason to need relief from their hard daily labors, who need to reaffirm life, to recover from the weariness, cares, and burdens of every day, to have their necessary pleasures ruined by the eternal chicanery of these pedants.

> We need more affirmation of life and less complaining! More morality, but less moralism! (Goebbels 1939, 385)[17]

Perhaps this can be Goebbels' view, and yet he was simply unable to reason himself into seeing the immorality of so much of what he did. Perhaps, as philosophers so often maintain, without relying on manifestly false or contradictory beliefs, he could avoid

---

17   Page reference is to the original text.

seeing what he did as wrong. However, I am not convinced that gross evil is so easily rationalized. We have a good idea how psychopaths manage to rationally maintain a coherent and amoral view of the world, but they are grossly deficient; too often do we suppose that those without such gross deficiencies follow their reason to much the same place.

In the above passage Baccarini also claims that "We blame and feel indignation for his wife for having killed their children [which we do], although this may perfectly follow from her evaluative standards." But again, I do not see why we must assume that it is quite so easy to have perspectives on the world that rationally imply the permissibility of these terrible actions (some accounts hold that Magda did not wish her children to grow up with the belief that their father was a criminal — does this tell us something?) This is especially so when we are talking about a mother murdering her children: that she had a goal, and even that she may have convinced herself that this would be a benefit to her children, does not show that a mother did not see that it was wrong to kill her children. In any event, it is very hard to know what was going on in Magda Goebbels mind in those last days in the bunker, but surely we cannot be confident in our moral reactions without knowing more about it. When a mother kills her own children, her mental state often does at least mitigate blame.

It should be stressed that *OPR* is concerned with the typical moral agent in our social morality, and the degree of care in deliberation that we expect of her. But, as I stress (*OPR*: 254), context matters — a lot. Those who occupy roles of political leadership are expected to be far, far, more deliberative than the "man in the street." What may have been "enough" deliberation for a German bus driver in 1939 would not be enough for a Minister. Someone in such a position is morally expected to think quite carefully about what he proposes. Now do we *really* think that Nazi politicians were models of good reasoners — that they reasoned well, albeit from bad commitments? Fascism, and even more Nazism, were premised on an attack on reason and an explicit devotion to myth and sheer will. Students of ideology typically see fascism as held together more by emotion and myth than by reasoned coherence. An ideology that constituted a revolt against reason strikes me, at any rate, as a distinctly inappropriate model of a rationally coherent immoral view.

What, then, of the final case — of parents who withhold medical attention from their children on religious grounds? Much depends here on what one thinks of religious conviction. Some secularists see faith as a sort of silly intellectual weakness, and so in this case would blame the religious person for immoral action and experience the full force of the moral emotions. "How could people be so stupid and immoral?" they might wonder. But this depends on a rather harsh view of people of faith. If one sees the parents as moved by a genuine but still mistaken conscience, then I think that one's moral emotions and judgments of blame must be greatly affected. As Adam Smith observed, "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." (1982 [1790], 176-7)[18]

5. Controversy and Public Reason Theories

I believe we are now in a position to consider Baccarini's claim that the forgoing analysis depends on a "a controversial philosophical position." There is one sense in which this is true, but not, I think, in the way that some critics have thought.

It is not true in the sense that the forgoing turned on the controversial view known as "reasons internalism," according to which reasons are only entities that, as it were, exist in the head. In *Justificatory Liberalism* I did endorse such a view; I concluded, however, that its analysis was indeed too restrictive. Note, however, that in section 4 (above) I did not employ the loctions "is a reason," "have a reason," or indeed, "a reason." The entire analysis was framed in terms of what a person can reason himself into endorsing or believing — it was about a mental activity, not an entity; the concern is whether people have deliberative routes to seeing their action as wrong. *OPR* takes no stand as to what "a reason" is. It takes no stand on whether there is a sort of mind-independent fact the matter that "φ is wrong," and whether this fact, in and of itself, is a reason not to φ. As one can infer from *Justificatory Liberalism*, I do not think these sorts of entities exist, but that is irrelevant to the analysis in *OPR*.

Of course there *is* a controversial philosophical doctrine in play here — *OPR*'s Strawsonian Claim. *OPR* certainly depends on it; as I have been stressing, it is what starts investigation into the participant perspective, and so pushes us into looking inside others' minds, and seeing the importance of the access condition for sustaining the reactive attitudes. *That* is the internalism in *OPR* that is controversial; it is not a thesis about reasons, but about our participant perspective in our moral practice. And of course it is philosophically controversial, for Baccarini and I are philosophers and we are now debating it.

If an account of public reason had to be uncontroversial in everything it said, it could only be a list of platitudes.[19] That is why, as I stressed at the outset, it is so important to keep in mind the different levels of theorizing in a public reason theory. *OPR* gives a philosophical analysis of our moral practice based on its Strawsonian Claim; although I believe it is an exceptionally well-justified claim as philosophical theses go, I certainly do not contend that it is uncontroversial. That would be crazy. Given this Strawsonian Claim, *OPR* argues that, to sustain our moral practices of blame, resentment and indignation among all the participants in society *S*, we are committed to moral claims of type *M* that are uncontroversial in group *S* the sense that:

> Claim I. We suppose that *M* would be endorsed (and so not disputed) by all members of *S* who reasoned with the level of care that our practice expects from them.

---

18 Van Schoelandt alerted me to the importance of this passage, which he discusses in his "Justification, Coercion, and the Place of Public Reason."

19   According to some, it would also have to be a platitude that they are platitudes, and, indeed platitude that it is platitude that they are platitudes…. See Gaus 2011, 227.

This is to be distinguished from a claim that must be rejected:

Claim II. We suppose that all members of *S* would interpret Claim I in the same way.

A sensible public reason account does not make the outlandish assertion that we always agree about the upshot of public reason. It would be absurd to think that theories of public reason imply that people who reason as well as they can will agree about most of morality. Think about the earlier example of the religious parents. When Alf interprets Claim I in his way, he believes the religious parents would see the wrongness of their action, and so he retains (perhaps, as Smith said, in a softer form), at least some of the reactive attitudes, and concludes they did wrong. Betty disagrees: as a religious person herself, she believes that the parents, reasoning as well as could be reasonable expected, would conclude that withholding medical care to their children was permissible. That both Alf and Betty accept the standard of public reason determines what they will argue about; it does not mean an end to argument.

It is also important to note that *OPR* does *not* accept:

Claim III.   If (*i*) Alf and Betty both endorse Claim I and:

(ii) Alf is claiming that Betty is morally required to φ, and she disagrees because they disagree about their interpretation of I, then

(*iii*) Betty's interpretation of Claim I is definitive unless Alf is sure beyond a reasonable doubt that she is wrong.

Baccarini is entirely right: it is "improbable on every interesting moral matter" (Baccarini 2013, 31) that Alf could so sure, and so it looks like one could never act in the face of disagreement. *Justificatory Liberalism* did indeed employ a "beyond reasonable doubt" standard at one point, but it is not part of *OPR*. Generally *Justificatory Liberalism* employed many epistemically idealized and stringent conditions. While I still think those idealizations led to important insights, an overriding aim of *OPR* was to think about public reason more realistically.

6. Models and Mental Experiments

I cannot really do justice to Miscevic's innovative analyses of thought experiments (TEs), and in particular political thought experiments. (PTEs) I can only touch on some of the important matters it brings to the fore.

Let me commence with Miscevic's characterization of thought experiments. A political thought experiment, he says, "is an episode of thinking that involves appeal to imagined, counterfactual situation in order to answer a moral-political question. On the most general level, features of political TEs are the following:

(a) Thought-experimental reasoning involves reasoning about a particular

set of social and political circumstances, which may be specified in more or less detail.

(b) The reasoner's mode of access to the scenario is via imagination rather than via observation.

(c) Contemplation of the scenario takes place with a specific purpose: forming judgment about some politically relevant theoretical proposal." (Miscevic 2013, 54)

I wish to consider more carefully (c), and the idea that the purpose of a political thought experiment is form a "judgment about some politically relevant theoretical proposal ." I think it may help to distinguish two interpretations of political thought experiments, depending on differences in the interpretation of (c). I begin with:

> *Mental Experiments*: The experiment sets out a counterfactual situation *C* in the imagination of Alf, which aims to elicit from Alf his judgment about *C*.

The Trolley Problem is, in my view, a quintessential Mental Experiment. The point of the experiment is to get a person to imagine the trolley coming down the track and to get the imaginer to choose which way to turn it. In this sense, Mental Experiments are inherently indexical: the result the experimenter aims at is her own judgment. So we might say:

> *The Indexicality of Mental Experiments*: Alf conducts an experiment to determine what judgments he will make about a counterfactual situation *C*.

We might be led to question the indexicality claim in two ways. When an objectivist like J. J. Thomson (1990, chap. 7) conducted a mental experiment such as the Trolley Problem, she was seeking an objective truth about moral principles. Thus we may think it was not *her* reaction that was important to her, but that it was the standard reaction that yields knowledge of moral truth. But the moral truth claim is not itself part of the experiment; it is what a theorist infers from the experimental results on the basis of a wider theory. However, an inference from an experiment is not itself part of the experiment (this point applies, I think, to some of the later stages of Miscevic's account of political thought experiments).

We might also get confused about the indexicality claim when we think about systematic experiments on the Trolley Problem by experimental philosophers and social psychologists (Greene, Sommerville, Nystrom, Darely, Cohen 2001). Here we have embedded experiments. The experimental philosopher is conducting a standard experiment about the outcomes of mental experiments. Indexicality applies to the mental experiment but not to the wider experiment of which it is part. The social psychologist is not seeking to elicit *his* judgment as to what will or did happen; his judgment is not the result of the experiment but quite the opposite: his judgment is to be formed by the result, and the

result concerns metal experiments by the subjects.

Contrast Mental Experiments thus conceived to a model. To fix ideas, think about a formal agent-based model run on a computer. Here we have some of the features of a mental experiment: the modeler sets up a counterfactual world (typically a very simple world, where many variables are left out, and the ones we specify relate according to simplified rules). In a way, the modeler sets up the entire world, for the rules, variables and parameters are of her own choosing. But in running the model she is not seeking to elicit her judgment about the world she has set up; the results of the model tell her about what her judgment about the world should be. Of course, her judgment does take a leading role in the evaluation of the model. If the results do not fit in with her overall theory or her other settled beliefs, she will think that it either missed something important, her rules are flawed, or the parameters ill-chosen. But just as we should not run together what one infers from an experiment with the experiment itself, we should not confused a model with its evaluation.

Now let us take the model out of the computer and put it in to the modeler's head. That is, she states her values, rules, and parameters, but she runs the model on her own hardware (brain) using her own software. This gives us:

> *Mental Models*: The model sets out a counterfactual situation *C* in the imagination of Betty. She sets up a world with rules, values and parameters, and seeks to run the model, predicting its outcomes.

Notice that indexicality does not apply to Mental Models; Betty is not interested in eliciting her judgment about the results of the model, she is interested in running it according to its rules, values and parameters, and making a judgment about what results are implicit in it. Moving the model from the computer to her brain did not change what she is seeking to do; it is a different way to go about it. Betty could move the model back to her laptop and run it again; she is interested in the results, not her reaction.

Confusing a Mental Model with a Mental Experiment can lead to serious misunderstanding of the point of the exercise. Some philosophers, I think, misinterpreted Rawls' original position, thinking that its aim was elicit our intuitive reaction to the choice situation. What we would choose is supposed to elicit one's intuition about justice; thus understood, one might interpret Rawls as a type of intuitionist. But this would be a mistake. Rawls, I believe, was seeking to set up a model of rational choice: running the model under those parameters is supposed to tell us what would be chosen. The model is then embedded in a theory which shows why this result is of moral importance — why choice under those parameters is of moral relevance. But again, that concerns the implications of the model, it is not part of the model itself.[20]

I finally come to my core point: in *OPR* the Deliberative Model is a Mental Model, not a Mental Experiment. The model sets up a social situation and seeks to derive what the

---

20 And, again, as with all models, if it gives us results that are inconsistent with other settled parts of our view, we are apt to think we got some of the rules, parameters or values wrong. This, as Rawls suggests, we would have to go back and revise the model.

result would be. This is an axiomatic model; everything needed to yield a result is in the rules (e.g., the Pareto-or-Indifference Rule), the values (e.g., the incomplete orderings of the Members of the Public), and parameters (e.g., mutual intelligibility, constraints on proposals). The aim is to not invoke a judgment about the counterfactual, simple, world, but to see what outcomes its rules, values, and parameters yield. It is, indeed, a pretty standard social choice model. And, as Miscevic nicely points out, all social choice models have two-tiered structure. We have a model of a group and its rules for choice, but the group is composed of model individuals, each with her own incomplete quasi-ordering: as he says, "the thinker imagines characters who deliberate together, and then s/he, the thinker, reaches conclusions from their imagined deliberation." (Miscevic 2013, 58-59)

However, we need to be careful not to interpret this as a Mental Experiment: the thinker's conclusion is not his reaction or intuitive judgment about what the Members of the Public should choose. It is, rather, his running of the model on his internal hardware/software configuration. Again, if the model yields odd results or violates some other commitment of the theory, the modeler may go back and try to find out what went wrong. But the Deliberative Model is not seeking to elicit a response by the reader, though she is invited to run the model on her own internal hardware/software configuration.

Miscevic is also correct that *OPR* "goes on to apply game theory to the imagined deliberation and interaction." (Miscevic 2013, 59) Thus the overall model has two stages: first we employ the Deliberative Model to yield a set of outputs and then we take those outputs as inputs into a simple iterated game model. But again, this is not a Mental Experiment. The aim is to take a simplified, counterfactual world that captures the crux of the justificatory problem and run it according to a set of rules, values and parameters to see what sort of outcomes we can achieve within the constraints. I hope there are important lessons to be learned from it about indeterminacy, path-dependency and ways to select from multiple equilibria. An especially important conclusion that emerges from the model is that to which Miscevic points: free and equal agents can converge on a single justified morality from a larger set without justifying a selection procedure. I believe that this is an important result.

To be sure, having developed a model and run it, a modeler always has to look back at the world to see if it enlightens us. Thus, as Miscevic points out, I am "keen on actual history" (Miscevic 2013, 59) — the abstract model gives way to discussion of the Magna Carta. But this is not part of the model; it is seeing how the model's results match up with historical experience. The model implies that history matters (since its results are path-dependent), so if we take it seriously, we must conclude "that contingent history can be justificatory." (Miscevic 2013, 59)

7. The Theory of Public Reason

These three thoughtful essays all raised fundamental questions about the theoretical commitments and apparatus employed in *OPR*. I have tried my best to advance the conversation as to why *OPR*'s focuses on social morality and public justification, why I do not think the nature of reasons is a significant issue, where the theory supposes agreement and where it does not, and the nature of the models it employs. Much more

could be said about these matters, and of course about all the important matters on which I have not touched. I apologize for these limitations.

I do not, though, apologize for *OPR*'s insistence that moral theory and political philosophy have gone astray. Philosophers all-too-often assume that they have moral insight that sets them apart from — and over — their fellow moral agents. *They* see the reasons that there really are, *they* see the true moral properties, *they* understand what justice demands about how society must be organized. Rather than seeking to understand our collective moral enterprise and to theorize about how it can function in a way that all can endorse, philosophers often claim to posses special *moral* insight that does not have to be confirmed by the reasoning of others. Rather than seeking to understand the great modern achievement of large-scale, free, diverse and cooperative societies and inquire how we can get them to better live up to their promise, many philosophers see their task as presenting us with detailed blueprints based of their own imaginations, insisting that existing structures that are constructed differently are *ipso facto* unjust. Theories of public reason reject this morally elitist project. This does not mean they are anarchistic holding that everyone can do as she chooses, that each person is the infallible judge of what is endorsable by her, or that she can simply decide to "veto" morality and so be free of its requirements. It does, though, mean that a morality for our democratic age takes seriously that people of good will in a diverse society have differing views of what constitutes an acceptable way of living together, and there is no priestly class who can dictate to the rest how we are to live together.

REFERENCES

Baccarini, E. 2013. Having a Reason and Distributive Justice in *The Order of Public Reason*. *European Journal of Analytic Philosophy* 9 (1): 25-51.

Baier, K. 1995. *The Rational and the Moral Order: The Social Roots of Reason and Morality*. La Salle, IL: Open Court.

Bistagnino, G. 2013. Gerald Gaus and the Task of Political Philosophy. *European Journal of Analytic Philosophy* 9 (1): 14-24.

Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.

Darwin, C. 2004 [1879]. *The Descent of Man*, second edition. New York: Penguin.

Estlund, D. 2011. Human Nature and the Limits (if Any) of Political Philosophy. *Philosophy & Public Affairs* 39: 207-235.

Gaus, G. 1990. *Value and Justification*. Cambridge: Cambridge University Press.

Gaus, G. 2013*a*. On Being Inside Social Morality and Seeing It. *Criminal Law and Philosophy*.

Gaus, G. 2013*b*. Why the Conventionalist Needs the Social Contract (and Vice Versa). *RMM (Rationality, Morality, and Markets)* 4: 71-87.

Gaus, G. 2013*c*. The Evolution, Evaluation, and Reform of Social Morality: A Hayekean

Analysis. In Hayek and the Modern Economy, ed. D. Levy and S. Peart. New York: Palgrave Macmillan: 59-88.

Gaus, G. forthcoming*a*. On the Appropriate Mode of Justifying a Public Moral Constitution. *The Harvard Review of Philosophy.*

Gaus, G. forthcoming*b*. Public Reason Liberalism. In *The Cambridge Companion to Liberalism*, ed. S. Wall and C. Kukathas. Cambridge: Cambridge University Press.

Gaus, G. forthcoming*c*. *The History of Public Reason in Political Philosophy*. Ed. P. Turner and G. Gaus. New York: Routledge.

Gaus, G. 2014. The Good, The Bad and the Ugly: Three Challenges to Public Reason. *Philosophical Studies:* DOI: 10.1007/s11098-013-0269-5.

Gauthier, D. 1991. Why Contractarianism? In *Contractarianism and Rational Choice*, ed. P. Vallentyne, 15-30. Cambridge: Cambridge University Press.

Gerald, G. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.

Goebbels, J. 1939. Mehr Moral, aber weniger Moralin! *Wetterleuchten*: 382-385. English translation *More Morality, Less Moralism!* translated by Bytwerk, R. 2000. *The German Propaganda Archive*. Accessed September 27, 2013. http://www.calvin.edu/academic/cas/gpa/goeb60.htm

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darely, J.M., and J.D. Cohen. 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293: 2105–8.

Hayek, F.A. 1973. *Rules and Order.* Chicago: University of Chicago Press.

Kant, I. 1999. *The Metaphysical Elements of Justice*, Second edition. Ed. and trans. J. Ladd. Indianapolis: Hackett.

Kitcher, P. 2011. *The Ethical Project.* Cambridge, MA: Harvard University Press.

Miscevic, N. 2013. A Hierarchy of Armchairs: Gerald Gaus on Political Thought Experiments. *European Journal of Analytic Philosophy* 9 (1): 52-63.

Shoemaker, D. 2011. Psychopathy, Responsibility, and the Moral/Conventional Distinction. *The Southern Journal of Philosophy* 49, *Special Supplemental*: 99-124.

Smith, A. 1982 [1790]. *The Theory of Moral Sentiments.* Indianapolis: Liberty Fund Inc. 176-7.

Strawson, P.F. 1961. Social Morality and Individual Ideal. *Philosophy* 36: 1–17.

Strawson, P.F. 1962. Freedom and Resentment. *Proceedings of the British Academy* 48: 187–211.

Thomson, J. J. 1990. The Trolley Problem. In *The Realm of Rights*, ed. J.J. Thomson, 176-202. Cambridge, MA: Harvard University Press.

The Department of Philosophy
College of Social & Behavioral Sciences
The University of Arizona
USA
ggaus@email.arizona.edu