Michael E. Bratman

Shared Agency: Replies to Ludwig, Pacherie, Petersson, Roth, and Smith

Abstract: These are replies to the discussions by Kirk Ludwig, Elizabeth Pacherie, Björn Petersson, Abraham Roth, and Thomas Smith of Michael E. Bratman, *Shared Agency: A Planning Theory of Acting Together* (Oxford University Press, 2014).

Let me begin by thanking each of the five contributors to this symposium.[1] In each case I have learned a very great deal in thinking through these insightful essays. And I do think that we are together – in a shared cooperative activity – making philosophical progress on these complex and important issues.

Reply to Kirk Ludwig

1. Ludwig focuses on the compressed account I offer of sufficient conditions for shared intention, an account that includes an intention condition, a belief condition, a persistence interdependence condition, and a common knowledge condition; and an account that also includes a mutual responsiveness condition for shared intentional activity. Ludwig wonders whether the intention condition by itself suffices

---

Michael E. Bratman, Department of Philosophy, Stanford University, Building 90 450, Serra Mall, Stanford, California 94305-2155, USA, Tel.: +650 723 2980, e-mail: bratman@stanford.edu

for shared intention, and whether, even given my strategy of focusing on sufficient conditions for shared intention, it is a mistake to add the belief condition, the persistence interdependence condition, and the common knowledge condition.

Suppose, he says, the intentions in the intention condition "are in fact satisfied. Then they J-ed together. Did they not thereby J together intentionally? ... if so, assuming that shared intention is what, when satisfied, results in collective activity, [the further three conditions] are not necessary for shared intention, nor necessary parts of a sufficient condition".[2]

In response, let me highlight why I was led to include these further three conditions. First, I wanted to show how my account, in appealing to each participant's intention that we J, could nevertheless acknowledge a settle condition on intention according to which in intending X one supposes that one's intention settles whether X. While I rejected an own-action condition on intention as insufficiently motivated, it did seem to me plausible that, as David Velleman had urged, intentions need to satisfy some such settle condition.[3] So I did not want my theory to depend on rejecting that condition. And Velleman asked how each of us, in a shared intention to J, could see her intention as settling whether we J and also see the other's intention as settling this matter.

My response was to explain how each of our intentions that we J could satisfy such a settle condition in circumstances in which it is believed that the intentions of each that we J are interdependent in their persistence and that taken together these intentions would be effective. In such circumstances each is in a position to believe that if she continues so to intend then so will the other, and that when they both so intend they will so act. (*Shared Agency*, 64–67)[4] The belief condition in the compressed thesis is motivated by this effort to provide a construction of shared intention that is in this way compatible with such a settle condition on the intentions of the participants. I then argued that in robust forms of shared intention the participants would not be making a mistake about the interdependence. And that led me to the interdependence condition itself.

It seems to me that Ludwig, in contrast, would advise that we give up the settle condition on the intentions involved in the construction of shared intention, though perhaps he would put in its place a weaker condition that there is some non-zero degree of credence in interdependence and efficacy. But my main point is that we do not need to pay that price in order to arrive at the kind of

---

2 Quotations from the essays to which I am responding are from the July 2014 manuscripts.

3 J. David Velleman, "How to Share an Intention," as reprinted in his *The Possibility of Practical Reason* (Oxford University Press, 2000).

4 I use these parenthetical references to refer to my *Shared Agency: A Planning Theory of Acting Together* (Oxford University Press, 2014).

construction I seek of robust forms of shared intention if we are willing to add to that construction these conditions of belief and interdependence. That said, we can still allow for weaker forms of shared intention that are primarily constituted by intentions that do not satisfy such a strong settle condition.

What about the condition that this is all out in the open, common knowledge? Ludwig also thinks this condition is not necessary for shared intention. In response, I would want to qualify Ludwig's comment that "shared intention is what, when satisfied, results in collective activity." The qualification I would offer is that shared intention is also what helps structure relevant forms of thinking together ("as a deliberative body," as Ludwig says) about how to proceed with the joint activity. This thinking together will normally involve the public recognition of the shared intention. And the normal way this will happen is when conditions sufficient for shared intention are out in the open. And that is why I included the common knowledge condition within the construction of shared intention.

In the background is a methodological concern. Ludwig warns that in focusing, as I do, on sufficient conditions for shared intention we need to be careful not to "include conditions not needed to constitute shared intention." So we should be "looking for a minimal sufficient condition for shared intention."

Now, the key constraint on my proposed construction is that it provide sufficient conditions for shared intentionality that are framed (with the possible exception of an appeal to common knowledge) in terms of resources available to the planning theory of individual agency, given knowledge of the minds and actions of others. This constraint is central to my defense of the continuity thesis. This strategy of sufficiency allows for the possibility of multiple constructions each of which respects this constraint, and each of which models a somewhat different form of shared intentionality. We need not decide which of these constructions gets at minimal sufficient conditions for shared intentionality in order to see the possibility of these various constructions as supporting the continuity thesis. Indeed, it is theoretically useful to see how our broadly individualistic resources enable us to model multiple, and in some cases increasingly robust forms of shared intentionality. What is central is that the conditions to which we appeal do indeed satisfy the cited constraint. So we will not want the construction directly to appeal to "mutual obligations among members" (though it can go on to note how shared intention normally engages relevant principles of moral obligation).

2. In essays that led up to my book I highlighted the condition that each participant in a shared intention intend that there be mesh in relevant sub-plans. I also highlighted the idea that the connection between shared intention and shared action normally involves mutual responsiveness in intention and action. In the book I took two further steps. I supposed that the intentions of each favored this condition of mutual responsiveness as a feature of the relevant connection

between thought and social action; and I supposed that the condition of intended mesh could be derived from these intentions in favor of mutual responsiveness. As Ludwig notes, however, I also acknowledged cases of prepackaged cooperation – for example, the synchronized divers (*Shared Agency*, 81) – in which at the time of action there is neither intended nor actual mutual responsiveness. But Ludwig notes that even in such cases we will want to include the condition that the participants intend that there be mesh in sub-plans. So "the requirement of meshing sub-plans should be put back in as a basic component."

I agree. The need for this adjustment does not challenge the claim in my book to have provided sufficient conditions for strong forms of sociality. In putting this condition back in we are not displacing the condition of intended mutual responsiveness. And it remains true that in standard cases in which there is intended mutual responsiveness the condition of intended mesh falls out of that condition. Nevertheless, I agree that in those prepackaged cases in which there is not intended mutual responsiveness we will want to put back in the requirement of intended mesh in sub-plans.

3. In appealing to the condition that each intends that we J by way of sub-plans of each that mesh, I helped myself to the strong idea that these would be intentions in favor of mesh all the way down. This fit with my strategy of sufficiency. But Ludwig worries that in many cases of shared intention the intentions of the participants favor only a weaker form of mesh in sub-plan, one that allows for "a kind of margin for error". And he asks whether we can more precisely "characterize the degree and kind of intended mesh characteristic of shared intention."

In response (and here I go beyond what I say in my book), let me try out a simple proposal. The proposal is that we appeal to the condition that we each intend that we J by way of sub-plans of each that *mesh to an extent that is sufficient to support the intended joint J-ing*. This allows for failures of mesh that do not undermine the joint J-ing.

Ludwig would say that this is overly simple since the gang members, in a version of the mafia example I discuss, might satisfy this condition without sharing an intention to go together to NYC. Perhaps each intends that they go to NYC, and each intends that the stage be set by the relevant intentions and meshing sub-plans of each. Yet each also intends, once the stage is set by meshing sub-plans, to throw the other into the trunk of the car and drive to NYC. Each intends that the sub-plans mesh sufficiently for the stage to be set for throwing the other into the trunk. So each intends that the sub-pans mesh to an extent that would be sufficient to support the joint activity of going to NYC together (in a sense that is neutral with respect to shared intentionality).

Note, however, that in the example neither gang member intends that the later stages of the joint activity of going to NYC involve responsiveness to the

intentions and actions of the other, who will be, it is supposed, in the car trunk. So neither intends that they go to NYC by way of on-going mutual responsiveness in intention and action. So long as we are careful to understand the intended mutual responsiveness condition to concern on-going mutual responsiveness, and not merely mutual responsiveness in some early phase of the activity, we can explain why these gang members do not share an intention to go together to NYC. This explanation appeals directly to the intended mutual responsiveness condition, rather than directly to a further qualification on the kind of mesh that is intended. But given this explanation, and putting aside cases of prepackaged cooperation, it seems plausible that we can retain our simple proposal concerning the kind of mesh that is intended.

Reply to Elisabeth Pacherie

My model of shared intentional action appeals to mutual responsiveness not just in intention but also in action. Pacherie supposes that this mutual responsiveness in action will "typically involve physical interactions and coordination between the agents" (section 4). She then notes that recent empirical research points to two conclusions about how in fact we achieve such physical interactions and coordination: "First, ... mutual responsiveness in action relies on a number of alignment and co-representation processes that are largely, if not always, automatic and involuntary and that involve, at least in part, sub-personal psychological mechanisms. Second, while these processes are crucial to shared agency, they are not needed to support individual action" (section 3). So there is reason to think that the step from the capacity for individual planning agency to the capacity for shared intentional agency involves distinctive sub-personal mechanisms of "alignment and co-representation" – of "construction of ... shared action spaces" (section 4) – mechanisms that go beyond those involved in individual planning agency. And that seems to challenge my continuity thesis.

The first point I'd like to make in response is that the phenomenon of human modest sociality goes beyond such cases of "physical interactions and coordination between the agents." There are many cases of shared activity in which the relevant mutual responsiveness in action is, rather, responsiveness in communicative action. Consider yours and my shared activity of trying together to make progress on a philosophical problem, or the shared activity of a group of architects who are together designing a building, or the shared activities of a search committee in trying to arrive at a hiring decision.

Further, even when the mutual responsiveness is primarily responsiveness in physical action it will many times exploit forms of person-level knowledge and communication rather than the kind of sub-personal "body glue" (section 4) that

Pacherie highlights. Suppose that we are painting the house together and that each of us is working in a different room. Each sends a text message to alert the other to her progress; the other adjusts and then communicates that. And so on. Or suppose the drivers of two cars in the process of making potentially conflicting turns coordinate their physical control of each car by way of the signaling systems of their cars.

Finally, and as Pacherie acknowledges, there can be shared intentional activity without mutual responsiveness in present action, as in the case of synchronized divers.

So the phenomenon of modest sociality goes beyond cases that involve "body glue". So Pacherie's reflections do not show that such sub-personal mechanisms of interpersonal physical coordination are necessary for modest sociality. So they do not show that we cannot provide sufficient conditions for modest sociality without appeal to these special kinds of sub-personal mechanisms. So her reflections are compatible with the continuity thesis.

The second point I'd like to make is that appeal to these sub-personal mechanisms does not affect the main debate between my approach and the approaches offered by Gilbert and Searle. The presence of the cited kinds of sub-personal capacities would not, I take it, ensure a capacity for a Gilbert-type joint commitment, or a Searle-type we-intention. (In a footnote Pacherie suggests that these sub-personal capacities might be included in what Searle calls the "Background"; but, in any case, that "Background" is not seen as ensuring we-intentions; nor does the use of this label show that these specific sub-personal capacities are necessary for shared intention.) Even given such sub-personal capacities we will need to ask whether, to get sufficient conditions for modest sociality, we need to appeal to such we-intentions or joint commitments. And here the planning theory claims that we do not need to introduce such a new element.

The debate here concerns the nature of overarching person-level structures of shared intention and the like, overarching structures that may sometimes be subserved by the special sub-personal mechanisms highlighted by Pacherie. And the main issue on the table is whether or not we can provide sufficient conditions for these overarching structures in terms of the theoretical resources provided by the planning theory of individual agency (given knowledge of the minds and actions of others).

Nevertheless, we can still ask about the truth and significance of the empirical claim that many cases of human modest sociality involve interpersonal physical interaction and coordination that is in fact explained by sub-personal mechanisms that "are not needed to support individual action". And here my main concern is that it does not follow from the (supposed) fact that the cited sub-personal mechanisms "are not needed to support individual action" that they are, in the relevant sense, discontinuous from those involved in individual action.

To see this, consider the appeal to a participant's intention that we J, in the plan-theoretic construction of a shared intention that we J. An intention that we J is "not needed" for individual action. But in introducing such an intention we are nevertheless drawing solely on conceptual, metaphysical and normative resources available within the planning theory of individual agency (given knowledge of the minds and actions of others). These resources include the attitude of intention, understood as a plan state, and the idea of our J-ing, understood in a way that is neutral with respect to shared intentionality. This idea of our J-ing does involve reference to others. But this is available to individual planning agents who know about the minds and actions of others. What this idea of our J-ing does not involve, at the ground level of the construction, is the idea of shared intentional activity. In this sense, the capacity to intend that we J is not a basic new capacity, over and above the capacities that are already involved in individual planning agency (given knowledge of the minds and actions of others). (And similarly concerning the capacity to intend that we J by way of your corresponding intention; and so on.)

As I see it, this contrasts with the relation between the capacity for planning agency and the capacity simply for present-goal-directed action. The complex cross-temporal intentional structures involved in the former seem unlikely to be constructible solely from materials from the latter even if those materials include ordinary knowledge of one's past and future.

So even if the "capacity for shared action space" (section 4) is "not needed to support individual action," it is a further question whether, in their appeal to this capacity, these empirical theories are appealing to a basic new capacity. Why not think of a shared action space as analogous to the representation of our J-ing, as it appears in the plan theoretic construction of shared intention? So understood, shared action space involves reference to others, but it need not draw on an irreducible idea of shared agency. Even if the capacity for shared action space is not needed for individual agency, why should we think that it is not, at bottom, constructible in a way that is analogous to the way in which, if I am right, the capacity for participating in a shared intention is?

I do not know the answer to this question, and welcome further reflections from Pacherie.

Reply to Björn Petersson

1. Petersson questions "whether there must be a *fundamental difference in kind* between Searle-style we-intentions and" the sort of intentions to which my theory appeals (28). And he seeks to answer No by providing "a way of understanding we-intentions that does not introduce a fundamentally different kind of entity to make room for collective intentions" (33).

Petersson calls the intentions to which my theory appeals "I-intentions". But that terminology is potentially misleading. I appeal, after all, to intentions, on the part of an individual, that *we* J. What is central to my theory is that these intentions that we J are, functionally and normatively, fully embeddable in the planning theory of the intentions of individuals, and that we can characterize these intentions without an irreducible appeal to the very idea of shared intentionality.

Suppose that you and I share an intention to smash the window. On my model of this shared intention, I intend that we smash it by way of relevant intentions of each, and you intend that we smash it by way of relevant intentions of each; and these intentions of each of us are intentions of the sort characterized by the planning theory. Drawing on ideas from François Recanati, Petersson proposes, as an alternative, that I have an intention in the "we-mode" (32–33) in favor of smashing the window, and so do you. The primary appeal to us is not in what is, strictly speaking, the content of each of our intentions; it is, rather, in the mode of the intentions of each. Nevertheless, the content of my intention taken together with its mode does determine what Recanati would call the "complete content"[5] which, as a first step, we can characterize as the proposition that we smash the window by way of each carrying out his or her relevant intention.

Petersson sees his appeal to such we-mode intentions as in the spirit of Searle's appeal to we-intentions. Is it? Well, recall the example of the pair of gang members each of whom intends that they go to NYC together, though each intends that this happen by throwing the other into the car trunk. Do these gang members have we-mode intentions in favor of their going to NYC together? I take it that even though the pair appears somehow in the intention of each, Petersson would say No. But why?

One answer might be that it is built into the we-mode that the resulting complete content involves shared intentional activity. And that is not how it is with the intentions of each gang member.

On this approach we are characterizing the we-mode in a way that draws directly on the very idea of shared intentionality. And this would be in the spirit of Searle's idea that there is no reduction to be had of this idea. But then intentions in the we-mode are fundamentally different from the intentions to which I appeal in my construction of shared intention. After all, according to my account, the intentions that appear in this construction need not involve, at the ground level, the very idea of shared intentionality.

---

5 François Recanati, *Perspectival Thought: A Plea for (Moderate) Relativism* (Oxford University Press, 2007), p. 5. Recanati contrasts "complete content" with "explicit content". Petersson's talk of content is, I take it, talk of what Recanati calls "explicit content".

An alternative explanation of why the gang members do not have the cited we-mode intentions – and I think this would be Petersson's explanation – is that the we-mode imposes the condition that the pair go to NYC by way of the relevant intentions of each; and neither intends that. The idea here is that we specify the conditions imposed by the we-mode not by appeal to the very idea of shared intentionality but rather by appeal to ideas that do not themselves involve that very idea – for example, an idea of the role of the intentions of each in the joint activity.

On this understanding of the we-mode, Petersson is faced with issues analogous to the issues I faced in my book concerning what else, over and above intending that we J, is needed to get sufficient conditions for shared intentionality. This is where I appeal to ideas of interlocking, intended mesh, intended mutual responsiveness, believed persistence interdependence, and so on. Petersson and I would be appealing to intentions of each that have similar complete contents – complete contents whose specification need not involve the very idea of shared intentionality. But Petersson would locate one of the determiners of that complete content in the mode of the attitude. So he would need to see the we-mode as having the theoretical burden of somehow bringing with it conditions concerning the role of the intentions of each, mesh in sub-plans, mutual responsiveness, and so on.

On this interpretation Petersson and I both accept the basic claim that we can give sufficient conditions for shared intention in terms of intentions with complete contents whose specification need not involve the very idea of shared intentionality. There remains the question whether or not to see this complete content as determined by a kind of content together with we-mode. But the main thrust of my account could accommodate either answer to this question, since in either case there is agreement about the basic claim. And neither answer would eliminate the gap between a theory like mine that appeals to intentions that need not involve the very idea of shared intentionality, and a theory like Searle's that requires appeal to intentions that do essentially involve the very idea of shared intentionality.

So on the first interpretation Petersson's we-mode intentions are indeed in the spirit of Searle's we-intentions. But on this reading there remains a basic difference between, on the one hand, we-mode intentions and Searle-type we-intentions and, on the other hand, the intentions to which I appeal. On the second interpretation the appeal to we-mode intentions does not capture the basic contrast that is at stake in the debate between my approach and Searle's. So it seems to me that neither interpretation supports the idea that there is not "a *fundamental difference in kind* between Searle-style we-intentions and" (28) the sort of intentions to which my theory appeals.

2. According to my theory, in the basic case of my intending that we J what I intend is that there be a coordinated concatenation of each of our individually inten-

tional activities that suffices for our J-ing. But what is this idea of our J-ing? A main claim I wanted to make was that (a) this idea, "while it does draw on ideas of individual intentional action, is neutral with respect to *shared* intentionality."(*Shared Agency*, 46) I also defended the further thought that (b) this concept of our J-ing "can be articulated using the conceptual resources of the planning theory of individual intentional agency." (*Shared Agency*, 46) In an earlier essay Petersson had suggested that the relevant idea of our J-ing, as it functions in a theory like mine, should be the idea that we J as group causal agent. And, while compatible with (a) (since the idea of a group causal agent need not bring with it the idea of shared intentionality), this seemed to me to be potentially in tension with (b).

My response was to explain how there can be basic cases of shared intention in which the intentions of each of the participants need not involve the idea of a group causal agent. Once such cases are available we can then identify sufficient conditions for a group causal agent, and we can allow that agents can then go on to incorporate that idea into the contents of their relevant intentions. (*Shared Agency*, 126) We can thereby accommodate group causal agents within the theory while holding onto (b) and without seeking a conceptual reduction of the very idea of a group causal agent to the resources of the planning theory of individual intentional agency.

In his present essay Petersson indicates that he would be sympathetic to such a conceptual reduction of the idea of a group causal agent to the resources of the planning theory of individual intentional agency. (In contrast, I did not want my theory to depend on the prospect of such a reduction, though I did not preclude it.) What is most important, however, is that he grants that we should allow "both ways of understanding 'We J' as it appears in the content of the participant's intention" (36). So there is, I take it, agreement that there are these two understandings of 'We J'; both satisfy (a); and both are possible for cases of shared intention (though Petersson is a bit worried here about relevant "linguistic intuitions" – 36). So it seems to me there is no significant difference between us here. Petersson does go on to say that this "difference between these two ways of understanding the intention condition... could e.g. make a difference when we apply the analyses to some of the moral or legal issues" (36) in this neighborhood. However, since my theory allows for both kinds of cases, it is fully compatible with – though it does not by itself entail – a proposal that this difference might matter in such ways.

Reply to Abraham Roth

1. Roth asks whether the conditions I supply "as sufficient for shared intention ... add up to guidance by interpersonal norms" (42). Roth is concerned that, if not,

this would show that my model does not succeed in its effort to provide sufficient conditions for shared intention and shared intentional action.

In the basic case, on my view, participants in a shared intention are guided by norms of individual plan rationality. I try to show that if there is this form of guidance, and the plan states of the participants are of the sort specified by my model, then there will be conformity to relevant norms of interpersonal plan consistency and coherence. There may also, in addition, be guidance by these interpersonal norms if the participants go on also to accept these. But, as I say, "in the basic case ... the norms accepted are, in the first instance, the rationality norms of individual planning agency."(*Shared Agency*, 33) What is crucial is the claim that this form of normative guidance will, given the distinctive contents and inter-relations of the plans of the participants, issue in conformity to relevant norms of social consistency and coherence.

Roth worries that this sort of individualistic normative guidance fails to ensure "the significance your intention is supposed to have for my practical thought" in cases of shared agency (43). To support this challenge he begins with an example of acting together with a "fungible partner" (43). I intend that we J in part by way of your intention that we J, but I am fully prepared to turn to a substitute for you "if you plan to implement your intention in a way at all incompatible with how I plan to implement mine" (43). The problem with this example, as Roth notes, is that it seems that I do not intend that we J "by way of relevant mutual responsiveness in sub-plan and action" and so by way of sub-plans that mesh. So Roth turns to an example of "playing chicken": "I intend that we ... J by means M1, and you intend that we J by some incompatible means M2. ... I hold onto my intention confident that you will back down ... simply because I don't change" (43). As I expect, you do indeed back down, our resulting sub-plans in fact mesh, and each of us is, according to Roth, "responsive to the intentions and actions of the other in ways that track the intended end of the joint action." (43, quoting *Shared Agency*, 79)

In this example, I fully intend that we J by way of meshing sub-plans; it is just that I am also fully confident that, so long as I stick to my rigid intention in favor of M1, these meshing sub-plans will, in the end, include M1. My intentions satisfy my side of the planning model of shared intention. Nevertheless, my inflexibility in sticking with M1 seems to Roth incompatible with the "the significance your intention is supposed to have for my practical thought" (43) in cases of shared agency.

What to say? A preliminary point is that, if Roth's example is to pose a clear challenge, we need to assume that it is out in the open that I rigidly intend that we go by way of M1 no matter what your earlier intention is. And if this is out in the open then you may well opt out of our shared intention. But suppose, despite this knowledge, you do not opt out. You are, as I expected, willing to tolerate and

work with my rigidity concerning M1. Perhaps you think "that is just MB being inflexible on such matters; I will go along with him, though M2 would be better." And we need also to suppose that your willingness to tolerate my inflexibility is a matter of rational functioning on your part.

A second preliminary point is that certain forms of coercion block cooperation, even in cases of shared intentional activity. (*Shared Agency* 101–102) If the success of my rigidity involves coercion on my part, then even if ours is a shared intentional activity it may fail to be a shared cooperative activity.

A third preliminary point is that ours can be a shared intentional activity even if it does not live up to relevant moral ideals of, for example, considerateness and respect.

These preliminaries in hand, the main point is that insofar as there is in the example a breakdown that blocks shared intention it is a breakdown in the kind of individualistic instrumental rationality that is presupposed by the plan-theoretic construction.

Begin by noting that rational intention normally involves minimal instrumental plasticity. Suppose I intend E and believe the world will in fact be such that a very specific M would be a necessary and sufficient means to E. I think that to achieve E I must thread the needle and proceed specifically by way of M. Nevertheless, I might still consider what to do if the world were unexpectedly to require that I instead perform means M', where M' differs from M only in minor ways. In intending E I will, if rational, normally not be unwilling to make any such minor adjustment in means if that were unexpectedly called for. (In *Shared Agency*, at 56–57, I drew on a similar idea in discussing the relation between intending that we J, and being disposed to help you play your necessary role in our J-ing.)

Granted, there may be cases in which any divergence from M, however small, would, as I see it, involve large costs that would outweigh whatever it is that speaks in favor of E. But it will be unusual for there to be no possible alternative means such that, if I newly came to believe that it instead was a necessary and sufficient means for E, I would, if rational, switch to intending that alternative means. The absence of any such potential alternative means would normally be a kind of fetish.

And now the point to note is that, in Roth's "playing chicken" example, my intention in favor of our J-ing by way of meshing sub-plans is, in relevant ways, not minimally instrumentally flexible, given my rigid adherence to M1. My failure to give your intentions appropriate "significance" involves a breakdown on my part in minimal instrumental plasticity. In special cases this might not be a rational breakdown; but normally it is. And if it is a rational breakdown then the planning theory need not say there is a shared intention to J. If, however, we are willing to say that this is a special case and that, despite my inflexibility,

my intention in favor of our J-ing by way of meshing sub-plans is functioning rationally, then – assuming all this is out in the open, and the other participant is rationally willing to tolerate and work with my inflexibility – we can see this as a case of shared intentional activity despite known but tolerated inflexibility. It can be a shared intentional activity even if it is not a shared cooperative activity and even if it does not conform to certain relevant moral ideals.

2. My model appeals to each participant's acceptance of norms of plan rationality. In explaining norm acceptance I refer to Allan Gibbard's development of this idea. But according to Gibbard a central feature of such norm acceptance is its role in social discussion. Roth's concern is that if that is also true of the norm acceptances to which I appeal, then there is a problematic circularity in my model of shared intention.

I indicated in my book that I thought we could explain the acceptance of norms of individual plan rationality without "essential appeal to the very idea of shared intentional agency." (*Shared Agency*, 16) Roth worries, however, that in the absence of some such appeal we will not be able to distinguish (as Gibbard insists we need to) guidance by a norm from being in the grip of a norm. But I think the way to make that distinction, while avoiding the circularity that worries Roth, is to draw on the approach to agential authority that I have discussed elsewhere. Roughly, I am guided by – and not merely in the grip of – norms that are functioning in their standard way and that have, because of their cross-temporal, Lockean roles in my individualistic psychic economy, and the absence of relevant conflict, authority to speak for me.[6]

3. Roth worries that my "invocation of the idea of norm acceptance … opens up the possibility that norms other than those of [individual] planning intention might be accepted, thus challenging the priority of [individual] planning agency over shared agency" (45).

On my view, however, the priority of individual planning agency does not depend on the absence of further kinds of norm acceptance. This priority is, rather, a theoretical conjecture: if we begin here, with these conceptual, metaphysical, and normative resources, we can construct robust forms of sociality. And the main effort of my book is to explain how to do that, and thereby to defuse the claim that there is a deep discontinuity between the individual and the social cases. Roth asks: "what sort of theoretical advantage is to be had by according primacy to planning agency and constructing shared agency from it." (46) And my answer is that it is the advantage of theoretical simplicity and explanatory power. Appeal to the core capacities of planning agency supports our understanding both of temporally extended

---

[6] See my "Three Theories of Self-Governance," as reprinted in my *Structures of Agency* (Oxford University Press, 2007), esp. p. 243–49.

individual agency and of central forms of social agency.[7] That is an aspect of the theoretical fecundity of planning agency.

This contrasts with Roth's idea that "both [individual] planning agency and shared agency are continuous with a form of agency that encompasses the possibility of acceptance" (46). What is key to the continuity I have highlighted between individual planning agency and social agency are the norms and structures of planning, not just the generic phenomenon of norm acceptance. This does not show that it is not possible to articulate a construction that has the form to which Roth alludes, one that sees norm acceptance as the underlying core, and individual and social agency as parallel developments from that core. But such an alternative construction would not be in a position to exploit planning structures in the way that my theory does in its account of shared agency. If we were presented with a theory that had this alternative structure we would then need to compare its theoretical virtues with those of the kind of planning theory I have developed.

4. Finally, Roth suggests that we should expand our conception of the characteristic functions of shared intention to include "*power sharing*": "Shared intention might have a function of distributing the authority or power to settle practical issues facing a group" (47). More specifically: the idea seems to be that it is a function of shared intention to support a balanced distribution of this "authority or power" (47) (which is to be distinguished from a balanced distribution of the benefits and burdens of a joint project). So "power imbalances take on a significance they would not otherwise have" (47). This is why Roth is skeptical that the guards and prisoners, in the example I discuss in my book (*Shared Agency*, 102), share an intention to build the bridge.

When I sought a way of specifying the characteristic roles of shared intention I sought parallels with individual intention. That led me to the idea that the central roles of shared intention include supporting interpersonal coordination of action and planning in some cases by way of supporting related bargaining and/or shared deliberation concerned with settling on relevant means and the like. When there is such coordination within the shared activity each participant rationally carries out her relevant plans for that shared activity in ways that support that shared activity. So the characteristic roles of shared intention include the role of accommodating the rational participation of each. And this is a kind of sharing of authority.

This last role of shared intention extends the parallel with individual intention. In planning for one's future one normally seeks to ensure that each of one's

---

7 On my view it also supports our understanding of our capacity for self-governance. See my *Structures of Agency*.

relevant future selves can rationally play their part in the temporally extended plan. The role of shared intention in accommodating the rational participation of each of the participants – and in this sense sharing authority – parallels this role of individual intention in accommodating the rational participation of each of the agent's future selves.

But this is not yet to appeal to a role of shared intention in support of a balanced distribution of power or authority in settling the main contours of the joint activity. Appeal to that role would seem to preclude shared intention in cases in which there is, in the background, significantly asymmetric influence over the main contours of the activity, yet familiar patterns of interdependent and interlocking intentions and of mutual responsiveness and pursuit of mesh in subplans. But, though I put such cases aside in my book, it seems to me – and I think Roth would agree – that shared activities with such background asymmetries in power or authority are not uncommon. Consider the shared activities of a parent and young child, or a teacher and a student, or a contractor and a sub-contractor. We can grant that shared intention has a characteristic role of accommodating the rational participation of each of the participants, while being skeptical that shared intention has a characteristic role of supporting a balanced distribution of power or authority in settling the main contours of the joint activity.

But could the guards and the prisoners really share an intention to build the bridge despite such significant differences in power? Well, despite these background power differences such a shared intention could exert local pressure, with respect to many details of the bridge building, in the direction of mutual responsiveness and flexibility, all in the direction of meshing, local sub-plans and the effective role of each in the joint activity. With respect to these local issues the intentions and plans of each can interlock and support mesh with the intentions and plans of all. Granted, the guards have the power to insist on certain features of the bridge building, for example: that the prisoners perform the most arduous labors. This poses a problem for the prisoners: do they participate in the shared intention or retreat, for example, simply to intentions to do what they need to do to avoid punishment? Some of the issues here will be similar to those I discussed in the first section of this reply to Roth. But such shared intentions in the face of significant inequalities of power nevertheless seem possible, though these differences in power may well block cooperativeness and/or fail to accord with relevant moral ideals.

Reply to Thomas Smith

Smith's main claim is that "BSI [Bratmanian shared intention] and GJC [Gilbertian joint commitment] are on a par with respect to 'deep continuity'. Both are

new kinds... Yet both can be accounted for by appeal only to existing kinds, as we can describe a sufficient condition that appeals only to existing kinds ... So ... the two postulates are on a par with respect to ontic parsimony" (56). So, Smith concludes, we should reject my claim that Gilbert's theory draws, in a way that my theory does not, on a basic new element in its account of our sociality.

Now, in introducing shared intentions I was indeed introducing a concept, and a "kind," that was not already part of the basic conceptual and metaphysical apparatus of the planning theory of individual agency. I provided a characterization of this "kind" that paralleled the functional and normative characterization I had provided for the intentions of individuals. As I had asked about individual intentions what roles they play in our lives and what norms are associated with those roles, I asked about shared intentions "what fundamental roles do they play in our lives, and what norms are associated with those roles?"(*Shared Agency*, 27) I said that "we should be struck by the analogues, in the shared case, of the coordinating, structuring, organizing, guiding, and settling roles of intention in the individual case." (*Shared Agency*, 27) And I highlighted that "as in the case of individual intention, we can expect the social roles characteristic of shared intention to be associated with characteristic norms of – in this case – social rationality." (*Shared Agency*, 28) In this way my initial characterization of this social concept and kind aimed to parallel my characterization of individual intention. I then argued – and this was the basic thesis in my book – that these social-norm-conforming roles would be realized by certain interconnected structures of the plan states of each participant. And that was the basis for my continuity thesis.

Smith sees Gilbert's theory as doing something similar. We begin with the social "kind" of joint commitment. We understand it in a way that parallels our understanding of the individual commitments involved in individual decisions. We then provide sufficient conditions for such joint commitments by appealing to relevant agreements between the participants. Smith notes that these proposed sufficient conditions do differ in a significant way. I seek conditions that constitute, as Smith says, "a persisting *presence* with a persisting power to guide action and planning." (55) In contrast, Gilbert, as interpreted by Smith, sees joint commitment as "an upshot state (like that of having promised) which consists in a prior act (of having jointly committed, by openly expressing readiness for joint commitment, etc.) and a persisting *absence* (of both recision and fulfilment of this commitment)." (55) So "while Bratman offers a *constitutively* sufficient condition, Gilbert offers a *causally* sufficient one" (55).

On this interpretation, Gilbert's causally sufficient condition for a joint commitment "is simply that the parties agree to do something" (55). Though such an inter-personal agreement is not part of my theory of individual planning agency, "it is scarcely a new *postulate*" (55).

So, on Smith's interpretation, both theories introduce a new social kind (shared intention, joint commitment), both understand it in a way that parallels an analogous individual kind (individual intention, individual commitment of the sort involved in individual decision), and both offer relevant sufficient conditions for the presence of that social kind, sufficient conditions that do not appeal to a "new postulate". And that is why Smith thinks the theories "are on a par with respect to ontic parsimony" (57).

In reply: Consider the purported capacity to create joint commitments by entering into interpersonal agreements that bring with them distinctive norms of interpersonal obligation. This capacity and these norms seem to go beyond the capacities and norms central to individual intentional agency. Perhaps appeal to interpersonal agreements is not a "new postulate" (55) in the sense that such agreements, and their associated norms, are familiar within our commonsense understanding of our sociality (which is not to preclude disagreement about what those substantive norms are). But such interpersonal agreements and norms seems to go beyond, and not to be constructible solely in terms of, what is involved in individual intentional agency.

In response Smith writes: "we may grant that Gilbert introduces norms that are discontinuous with those of individual agency *as theorized by Bratman*. But they are not discontinuous with those of individual agency *as theorized by Gilbert*. For her, any creator(s) of a practical commitment, be they singular or plural, owe themselves its fulfilment (or failing that its recision)" (52). But we need to distinguish three ideas:

1. An individual decision creates a practical commitment that involves normative pressure for sticking with that decision.
2. There is a structural parallel between the capacity for individual decision, as in (1), and the capacity for participating in the creation of plural commitments that involve distinctive inter-personal obligations.
3. The capacity to participate in the creation of such plural commitments does not go beyond capacities that are involved in individual agency, including the capacity for individual decisions, as in (1).

And my concern is that even if we were to accept (1) and (2), we still would not have adequate grounds for (3). This is because, even granting the parallel in (2), we would still need to acknowledge that plural commitments involve inter-personal normative ties that seem to go beyond and not be constructible solely in terms of the individualistic normativity in (1). But it is (3) that is needed to support Smith's claim that the theories "are on a par with respect to ontic parsimony" (57).

Perhaps (though I would be skeptical) there is an argument that there could not be individual agents who had the capacity for individual decision as in (1) but did not have the capacity to participate in the creation of Gilbertian plural commitments. But to my knowledge this is not an argument that Gilbert makes. And in the absence of some such argument we have not been given reason to accept (3).

In any case, as I argue in my book, so long as the concept of shared intention is the concept of that which is apt to explain relevant joint activity, it seems false that such agreements are sufficient for shared intention. As I say: "people can insincerely agree to, say, plow the fields together, even though each participant fully intends not to act in accord with that agreement."(*Shared Agency*, 116; and see 112) In such a case of insincere agreement, while there may well be a mutual obligation to plow, there is not the sort of shared intention that is apt to explain a joint activity of plowing the fields. What there is, rather, is an intentional structure that is apt to explain why the fields are not plowed. But Gilbert supposes that if there is joint commitment there is shared intention. So the cited agreements are not sufficient for joint commitment, so understood.

This is related to Smith's thought that we need to allow for the possibility of "rationally *complying* with, but not *embracing* a shared commitment." One might rationally comply with one's side of a mutual obligation one regrets and in that sense does not embrace. But, asks Smith, what if the shared commitment in question is the kind of intentional structure in terms of which I model shared intention? Can one rationally comply with a shared intention in which one participates, and yet not embrace it? Yes, since one might rationally intend that we J without any enthusiasm about the prospect of our J-ing. But there will be kinds of failure of embrace that will block shared intention. In particular, if (as in our case of insincere agreement) one intends not to play one's necessary part in J then one will not be in a position rationally to intend that we J, and so will not be in a position rationally to participate in a shared intention to J. So, assuming rationality, this kind of lack of embrace blocks shared intention. It need not, however, block relevant (as Smith says, "quasi-contractual"; 57) mutual obligations; and one may go on rationally to comply with those obligations, though this might require some change in intention.