Notes:
Michael Bratman, *A Theory of Shared Agency* (Yale Draft, June 2011)

Stephen A. Butterfill
<s.butterfill@warwick.ac.uk>

October 16, 2011

Contents

[These notes are not my own work—I'm drawing on the Yale discussion.]

1.   What is shared agency?

It seems to matter that the reader has a handle on what shared intention and modest sociality are because this defines the book's aim: 'my primary concern is not with our pre-analytic talk but with shared intention as a central element in the explanation of activities involved in ... modest sociality' (p. 101).

I take it that the term 'shared' doesn't by itself fix a topic, and that 'shared intention' is a term of art (but see section 5 on page 8: maybe we have been too quick to rule out the idea that shared agency involves shared intention in just the sense of sharing in which two people can share a name). By contrast, the manuscript talks about 'the shared-ness of the activity' (p. 47) and asks whether something 'suffice[s] for 'shared-ness" (p. 125), which I read as presupposing that there is some target

aspect, the shared-ness, out there to be captured. This seems mysterious. After all, the three legs of a tripod share the work of keeping it upright; but their activity (that of supporting the structure) does not feature the relevant kind of shared-ness. So what kind of shared-ness is at issue here?

Relatedly, 'the approach … aims to provide a substantive account of that in which the shared-ness of shared intention consists' (p. 144). For reasons given below (section 5 on page 8), I wonder whether the sharedness of our shared intention consists just in us each intending that we ϕ. If so, the shared-ness of shared intention is not terribly deep; it's approximately the sort of shared-ness exemplified by two people who share a name. But even if this were correct—even if our sharing an intention were just a matter of us each intending that we ϕ—it seems clear that the true aims of Bratman's account would not be adequately met by describing the sharing.

And 'sometimes deception and coercion between the participants in an activity clearly block the shared-ness of the activity' (p. 47). While it seems clear (from the discussion) that deception and coercion sometimes undermine coordination of planning, it seems hard to know what might determine whether shared-ness were blocked.

In addition to supposing that appeal to sharing doesn't help, I also take it that the contrast cases don't serve to fix a topic. Take the contrast between walking together and walking besides the stranger. If the question is how these differ, there are lots of possible answers including some which appeal to emotion and phenomenology. A quick way with these examples can be had by appeal merely to goal-directed action. In walking together, there will be some outcome to which each of our actions is directed. For instance, it may be that our both arriving at the top of 5th Avenue is an outcome to which each of our actions is directed. So neither of us can succeed unless we both succeed. By contrast, in the case of the stranger, there is no goal to which both of our actions are directed. My getting to the top of 5th Avenue is an outcome to which my actions are directed; Stranger's getting there is an outcome to which her actions are directed. But there is no outcome to which an action of mine and an action of Stranger's are both directed. So we can distinguish walking together from Stranger by noting that in walking together only there is a single outcome to which actions by each of us are directed. (The same applies to Searle's contrast case involving park visitors.) Of course, this way of distinguishing the contrast cases does not introduce any interesting notion of shared agency. The mere fact that there is an outcome to which both of our actions is directed does not even imply that our actions are coordinated.

A simpler approach would be to say that the topic is coordination of

action and planning; or to treat the roles of shared intention as topic-defining stipulations rather than discoveries. Of course this might affect parts of the argument (perhaps because not all potential opponents would agree on this specification of the topic). But it has the advantage of enabling a clear distinction between premises and discoveries.

To illustrate this advantage, consider the claims that 'in shared intention each participant is committed to treating the other participants .. as ... intentional co-participants in the shared activity' (p. 59) and 'in shared intention the fact of shared intention will normally be out in the open' (p. 71). If these are discoveries, how are they known? If we accept that shared intention functions to coordination action and planning, the justification for these claims seems straightforward given what we learn from the book. But neither claim seems obviously true if we rely on the examples to fix what shared agency is. (Suppose that each of two people lifting a heavy sofa together falsely but sincerely believes that the other is a non-intentional, single-purpose, sofa-lifting robot; intuitions may differ but it doesn't seem obvious that their having these beliefs is incompatible with their having a shared intention of some kind, although this would arguably be incompatible with both claims. The issue here isn't whether these two people have a shared intention; it's how we could know either way unless we anchor shared intention using its role in coordinating plans.)

2. minor things

p. 33: 'the apparent challenge .... we are interested in our shared agency, and this is shared agency whose participants are ... planning agents'. — Although I'm not sure what the apparent challenge is, I'm also suspicious of the reply. We are planning agents but we are not only planning agents, so our shared agency may have aspects entirely independent of our capacities for planning (see section 4 on page 5).

p. 85: 'we will at least want to add ... relevant beliefs ... about effectiveness'. — I haven't managed to figure out why. I understand that (i)–(iii) must *be* coherent. I suppose this means that an agent who believed the negation of the belief content in (iv) would be in trouble, and perhaps the same is true of an agent who merely had, on balance, evidence in favour of this. But I don't see why agents must have the belief described in (iv). And the point that the account provides merely sufficient conditions doesn't seem to work here. This is partly because the account is also supposed to be psychologically realistic (in which case less is surely better than more). And partly because if the substantial account failed to apply in cases where condition (iv) did not obtain but some such

cases did involve shared intention, then it would weaken the presumption generated by the argument against the continuity thesis. (Here I'm supposing that the more cases of shared intention are explained by an account meeting the continuity requirement, the stronger the presumption in favour of continuity.)

p. 87: 'it is independently plausible that one characteristic feature of shared intention and modest sociality is that there is something like this interdependence between the participants' — why is this plausible?

p. 119: 'the complex content of the intentions ... may only be implicit' — as I said in discussion, I think this is a reasonable reply to the objection that ordinary agents with a shared intention cannot always *articulate* the intentions involved in having a shared intention. But I don't think it is a good reply to the (potentially more interesting) objection that meeting the sufficient conditions for shared intentions may be cognitively demanding in the sense that getting into such a state may require some time, may consume working memory and may place demands on executive function (so that someone who had to act very quickly, or who had limited working memory might be unable to form a shared intention).[1] However, I think that making this objection convincing would require providing empirical support for the claim that meeting these conditions is cognitively demanding for adult humans (***provide Geurts reference where he rejects objections to Gricean structures based on cognitive demands).

3. fn. 254, p. 162

I read this note as saying that there are reasons for thinking that a claim Ludwig (2007) makes is false, or at least pointing to a contrast between Ludwig and Bratman. I don't think the footnote entirely fair to Ludwig, however. (I mention this because, despite many references to other work, it's the only place I where I thought Bratman might be being less charitable than he could be.) In Bratman's discussion (in Chapter 7), there are (it seems to me) two senses of agency in play. In any shared intentional action, each individual agent is an $agent_1$ of an action. In addition there is a 'group causal agent' which is an $agent_2$ of an action. I take it that *agent*$_1$ will be explained in terms of intention plus motivational potential, whereas *agent*$_2$ is a less basic notion in the sense that it is best explained as an attenuation of agency$_1$. I think it's reasonable, in

---

[1] Compare Knoblich & Sebanz (2008, p. 2022): 'the contribution of lower-level processes to social interaction has hardly been considered. This has led philosophers to postulate complex intentional structures that often seem to be beyond human cognitive ability in real-time social interactions.'

this context, to take Ludwig to be using the term 'plural agent' to mean something which is an agent$_1$ of a joint action. And I don't think that Bratman has given an argument against the claim that only a subject with beliefs could be an agent$_1$ of an action. (Indeed, Bratman would probably not want to contest this claim.) So I don't think that Bratman is rejecting any claim which Ludwig makes.

I also suspect that Ludwig's paper may support an objection to Bratman's claim (on p. 159) that 'if 1. is true in the way envisaged by the basic thesis then there is this group causal agent and that group agent can in fact be the referent of 'we' in 1.' If Ludwig is right (and I haven't seen a good objection to his analysis), 'we' should be treated as 'as in effect a quantified noun phrase' (Ludwig 2007, p. 364). To expand slightly: the idea is (i) there are contexts in which sentences involving 'we' like 1. (on p. 159) are true where there is no causal group agent; (ii) in those cases, the best semantic theory offered so far treats 'we' as a quantified noun phrase; (iii) we can give a uniform account by applying the same semantic theory to 1. (on p. 159); (iv) a uniform account is better than a non-uniform account. If this is right, it is not true that a 'group agent can ... be the referent of 'we''.

4.   Shared agency which does not engage planning abilities

Suppose there are forms of shared agency that do not engage any of the agents' capacities for planning and which are entirely independent of their capacities (if any) for planning. Call this *non-planning* shared agency.

(In the discussion we seemed to fix on labelling these 'proto'. I think this is a loaded label; it might easily be taken to suggest that non-proto phenomena have some kind of conceptual priority, or that the value of the proto phenomena consists in part in their being developmental or evolutionary pre-cursors of the non-proto phenomena. In the long run everything is proto or it's the end of the world.)

One argument for the existence of non-planning shared agency might lean on developmental research.[2] But once we concede that any non-planning agency exists, it seems possible that even agents with planning abilities sometimes engage in actions exemplifying these less demanding

---

[2] From around 18-months children can coordinate their actions with others sufficiently well to do engage in activities directed to novel goals (such as bouncing a cube on a large trampoline) with another agent (Warneken, Chen & Tomasello 2006) where their actions are likely to be voluntary not only with respect to the outcome but also with respect to whether they are acting with another agent as contrasted with acting in parallel with another agent (Gräfenhain, Behne, Carpenter & Tomasello 2009).

forms of shared agency—perhaps where they lack time, energy or inclination to plan, or perhaps (relatedly) in performing actions which are components of full-blown shared intentional actions.

A related argument for the existence of non-planning forms of shared agency would be that we have a plausible account of it which satisfies a version of the continuity requirement. At least I try to provide such an account (Butterfill 2011a,b).

I suggest the existence of non-planning shared agency would raise five issues for Bratman's argument. None are fatal; in fact most might be seen as favouring the general approach.

The first issue concerns necessary conditions for shared agency. We might take Bratman to claim that all shared agency involves states or dispositions whose functional roles include coordination of planning (so not only coordination of action).[3] If we take the pre-theoretical notion of shared agency to be anchored by a series of cases such as running a give-and-go (and perhaps equally by developmental cases such as jointly bouncing a block on a large trampoline), then it seems the claim cannot be taken for granted. But what is the argument for it?

The second issue concerns the argument for the continuity thesis, which appeals to Ockham's razor. To establish the continuity thesis, Bratman need only show that there is one model which applies to all cases of shared agency and meets the continuity requirement. However, suppose that there are cases of shared agency that Bratman's model fails to characterise. Then further argument is needed. (While I'm hopeful that the further arguments are available, I do think this is a possible line of objection for a proponent of a Searle-like discontinuity to push. This results in a kind of symmetry between Gilbert-style and Searle-inspired objections: the former says Bratman's hasn't sufficiently focused in on the genuine phenomena, the latter says his focus is too narrow.)

Third, insofar as the planning model of shared agency is intended to offer a realistic psychological description (in this respect Bratman seems to be more concerned with scientific investigation than Grice was in theorising about meaning), the existence of shared agency without planning may complicate the model. For suppose that we have a notion of shared agency that does not engage planning capacities and therefore does not involve shared intention (at least not as characterised by Bratman). Then without circularity we can appeal to this notion in characterising the contents of intentions, including an intention that we J. One might think that this is not a major issue by analogy with the case of individual action: while many including Bratman allow that there may be actions which

---

[3]   I say this based more on the discussion than the manuscript. As far as I can tell, this is not an explicit commitment, but it may be a premise required for the argument for the continuity thesis (see the second issue below).

are goal-directed but do not involve full-blown intentions, few theories of intention draw on the resources provided by a theory of more basic kinds of goal-directed action. However it is also possible that this is a weakness of theories of intention, although perhaps not one that will change the theoretical landscape in the sense of undermining arguments for the irreducibility of intention. In both individual and shared intention, it may be that a fully adequate theory should explain how intentions interface with other forms of cognition necessary to get from the intention to the bodily movements which ultimately realise it (where the intention's realisation requires bodily movements). And this explanation my depend on exactly how the contents of intentions are characterised.[4]

Fourth, the existence of non-planning shared agency may complicate the details of Bratman's argument. For suppose that in intending that we J, we each characterise our J-ing not as a bare cooperatively neutral action but as an action involving non-planning shared agency. Then it it is plausible that the intention that we J will already exclude mafia cases. For in these cases there is no coordination of action, or too little coordination of action for them to count as cases of non-planning shared agency.

Fifth, acknowledging the existence of non-planning shared agency makes clearer the value of the notions provided by the planning theory. The reason we value shared intention is not primarily that it enables us to coordinate our actions but that it enables us to coordinate our plans. This issue becomes vivid when we imagine agents who have no planning abilities. They may enjoy some form of shared agency, but they will only succeed insofar as either there is no need for their subplans to be coordinated (so in relatively simple cases, not typically actions which take any length of time), or else insofar as their environment provides

---

[4] Vesper, Butterfill, Knoblich & Sebanz (2010) raise this sort of issue for shared agency, and the corresponding issue for individual action is the topic of a paper I'm working on with Corrado Sinigaglia. To make this issue pressing we might ask how intentions could result in bodily movements. In the case of humans, intentions typically or always affect bodily movement through motor cognition. Motor cognition appears to involve relatively rich representations of action (e.g. representations in terms of grasping or reaching as directed to a particular target object which can be realised in many different ways in different situations, not only representations of particular muscle contractions or movements). So for my intention that I grasp a cup (say) to succeed, that intention must set a standard of success for motor cognition. Given that intentions and motor representations of action employ different representational formats—arguably one is propositional whereas the other is not—the possibility that motor representations of action are somehow related to the contents of intentions, perhaps by means of some demonstrative element, could be essential for understanding how intentions interface with motor cognition. And there are also parallels here concerning worries about circularity in specifying the contents of intentions.

for the coordination of their planning. To put the idea crudely, the move from swarm-like group behaviour to non-planning shared agency allows coordination of actions directed to potentially novel goals; and the further move from non-planning to plan-based shared agency allows for coordination of potentially novel plans and subplans.

5.   A deflationary (but friendly) alternative

[In conversation you suggested writing this up (in the spirit of mapping the territory), which I'm now planning to do]

Intention is characterised in part by norms such as agglomeration (according to which it is not rational to have several intentions unless it is rational to have a single intention agglomerating them all). On Bratman's view, the norms characteristic of intention concern for the temporal coordination of an individual agent's planning. But they do not concern for the interpersonal coordination of two or more agents' plans; this is the concern of further norms, norms of shared intention. However, being struck by the parallels Bratman draws between inter- and intra-personal norms, it may seem natural to suppose that the norms characteristic of intention, properly understood, already incorporate interpersonal cases. So take agglomeration. Consider how it might apply to one or more agents with several intentions, $p_1, ... p_n$. Suppose that each p_i specifies agents non-indexically (so, for example, Bratman intends that Bratman ϕ and Facundo intends that Facudo ψ, and Bratman and Facundo each intend that Bratman and Facundo χ). Agglomeration says it is not rational for this agent or these agents to have these intentions unless it is rational for each agent to intend that p_1 and ... p_n. So, in the above example, it must be rational for Bratman and Facundo to each have a single intention with a content specifying that Bratman ϕ and Facundo ψ and Bratman and Facundo χ. (Of course this way of formulating the norm may make it too strong; it might fail to be rational for Bratman to have such an intention on the grounds that intentions settle what is to be done and Bratman is not in a position to settle this. The guiding idea would be that, where agents intend that they ϕ, for any agent whether it is rational for them to have this intention depends on each agent's conformity with the norm.) If we suppose that intention is characterised by norms which already incorporate normative requirements on shared planning agency, if these norms entail the norms Bratman associates with shared intention, then it seems plausible that our shared intention could consist just in each of us intending that we ϕ. To suppose that some more complex content, or common knowledge, was needed in order to characterise shared intention would be a mistake. It would be like attempting to characterise

what it is for an individual to intend that she ϕ by saying that she has to intend, not only that she ϕ, but that she ϕ in accordance with and because of her intention that she ϕ. Of course this is not to deny that shared intention often relies on common knowledge, for it seems likely that conforming to the norms of shared planning agency does often require intentions to be common knowledge. However, unless this is a *necessary* condition on shared intention, it may be that explaining what shared intention is (as opposed to demonstrating the plausibility of the account) need not involve appeal to common knowledge.

So on this view (if it works) there would be no multiple realisability. For us to share an intention that we ϕ is for us each to intend that we ϕ.[5] While our so intending may often be rational only when our intentions are common knowledge, this is no part of the account of what shared intention is.

References

Butterfill, S. (2011a). Joint action and development. *Philosophical Quarterly*, *forthcoming*.

Butterfill, S. (submitted 2011b). What is joint action? a modestly deflationary account.

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, *45*(5), 1430–1443.

Knoblich, G. & Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B*, *363*, 2021–2031.

Ludwig, K. (2007). Collective intentional behavior from the standpoint of semantics. *Nous*, *41*(3), 355–393.

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8-9), 998–1003.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development*, *77*(3), 640–663.

---

[5] Contrast (p. 144): 'shared-ness of shared intention consists, roughly, in the interlocking and interdependence of planning attitudes of each, planning attitudes whose contents favor the joint activity and the meshing roles of both, all in a context of common knowledge'