

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

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

October 16, 2011

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

1. 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 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’.

2. 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. 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.¹ 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 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.

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

¹ 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).

not only coordination of action).² 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.

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 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 may depend on exactly how the contents of intentions are characterised.³

² 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).

³ 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

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 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 even 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 subplans. 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 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.

3. What is shared agency?

4. A deflationary (but friendly) alternative

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 provide for the temporal coordination of an individual agent's planning. But they do not provide for the interpersonal coordination of two or more agents' plans. However, being struck by the parallels between inter- and intra-personal norms, it

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.

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, \dots, p_n . Suppose that each p_i specifies agents non-indexically (so Bratman intends that Bratman ϕ s and Facundo intends that Facundo ψ s, and Bratman and Facundo each intend that Bratman and Facundo χ s). Then agglomeration says it is not rational for her or them to have these intentions unless it is rational for each agent to intend that the conjunction of these intentions. If we suppose that intention is characterised by norms which already incorporate normative requirements on shared planning agency, then it seems plausible that our shared intention consists 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 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 is unclear why explaining what shared intention is (as opposed to demonstrating the plausibility of the account) should 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 . 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.
- 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.