Notes:
Michael Bratman, *A Theory of Shared Agency*
(Yale Draft, 6 September 2011)

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

October 19, 2011

Contents

***

1. What is shared agency?

The book's aim is defined in terms of shared intention and modest sociality: '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). What are those?

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 3 on page 6: 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?

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.

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 3 on page 6), I wonder whether the shared-ness of our shared intention consists just in us each intending that we ϕ. If so, the shared-ness of shared intention would not be terribly deep; it's approximately the sort of shared-ness exemplified by two people who share a name. Yet if this were correct—if our sharing an intention were just a matter of us each intending that we ϕ—it seems that (presentation and details aside) Bratman's account would not be profoundly affected. So because I wonder whether characterising shared-ness is a genuinely deep project, I also wonder whether the aim is really to characterise shared-ness.

In addition to supposing that appeal to sharing doesn't help, I also take it that the standard 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 more clearly distinguishing what is established by argument from claims whose justification is more intuitive.

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. 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.[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 consideration in favour of 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).[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. (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

[1] From around 18-months children can coordinate their actions with others sufficiently well to 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). Yet there is some evidence (not decisive but substantial) against supposing that they are able to track others' intentions (I describe this briefly in Butterfill (2011a)).

[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).

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 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.[3]

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

[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 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.

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 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. 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 unless someone offers a devastating objection first. I think the general idea is very much in the spirit of Bratman's account, and, presentation aside, doesn't change very much. In effect, it amounts to going one step further than the continuity thesis. If not all joint action involves shared intention, the idea is also an extension of the view that joint action is just action (which I try to defend in Butterfill (2011b)).]

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. And there is the tricky issue of determining exactly which intentions Agglomeration* applies to—if we aren't careful about social networks we could end up with an implausibly strong requirement. But the guiding idea is just to think of meshing as a requirement on intention generally, not one that applies only to shared intentions.) If we suppose that intention is characterised by norms which already incorporate the normative requirements 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.[4] 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.

[4] 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'

4. 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''.

5. Psychological demands

p. 118-9: 'couldn't there be agents—four-year old humans, perhaps—who engage in a form of modest sociality but for whom such complexity is not yet psychologically available? ... [1] the complex content of the intentions ... may only be implicit ... the intentions cited in the intention condition are not as psychologically demanding as they may at first seem ... [2] certain less conceptually sophisticated psychological phenomena might in certain circumstances substitute for these more conceptually demanding attitudes.'

As I said in discussion, I think that Bratman's first point, [1], 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 there are two potentially more interesting lines of objection (corresponding to two ways of understanding your phrase 'psychologically demanding'). The first line of objection concerns conceptual sophistication, the second cognitive demands.

The first line of objection is that (i) a certain degree of conceptual sophistication is required to track others' intentions (where intentions are understood as elements of plans, of course; this wouldn't apply if we construe any goal representation as an intention); and (ii) not all agents capable of modest sociality have the required degree of conceptual sophistication.[5]

The second line of objection is that (i) 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); and (ii) in some cases agents who do are unable to meet these cognitive demands (either because they have limited cognitive resources or because they are simultaneously engaged in other activities (such as counting backwards) or because they are suffering some form of deprivation or temporary lesion) are nevertheless able to act in ways characteristic of modest sociality.

In reply to both objections, one part of Bratman's reply is (in effect) that the truth of their second premises has not been established. I agree, of course, that the truth of these premises is partly an empirical matter. While I don't know of decisive evidence in either case, I do think there is enough empirical evidence to make these lines of objection worth considering.[6]

[5] I don't understand clearly enough what determines whether an episode involves modest sociality to know how to make this line of argument work, but I gather from the discussion that Bratman is inclined to suppose that children's stacking blocks together (or bouncing a block on a large trampoline together) as exemplifying modest sociality. This would motivate a focus on one- and two-year-old humans rather than four-year-olds (see footnote 1 on page 4).

[6] See below on the second line of objection (cognitively demanding). On the first line (conceptual sophistication), while relatively little research has considered abilities to track intention there is some evidence that children of this age have difficulty understanding intentions (Astington 1991; Astington & Gopnik 1991). A range of researchers have argued that infants form expectations about goal-directed activity (Csibra 2008; Gergely, Nadasky, Csibra & Biro 1995; Woodward 1998; Woodward

Does Bratman's first point, [1], suggest a good reply either of these objections? I don't think this carries much weight. Put roughly, this is not a good reply because appealing to implicit states does not amount to explaining how something which appears to be effortful can be accomplished without effort, nor how something which demands conceptual sophistication can be achieved by those with limited intellectual powers. Appealing to dispositions only pushes the issue back one step to the question of what grounds these dispositions. If, as seems plausible, my dispositions to track others' knowledge of my intentions concerning their intentions are grounded by representations of their knowledge of my intentions concerning their intentions, no progress has been made; certainly we lack any detailed non-representational account of how such representations might be grounded.[7]

But perhaps this is too quick. Bratman's first point, [1], might be taken as somehow leaning on the views of those who think that mental state concepts are innate or at least appear early in development (e.g. Baillargeon, Scott & He 2010; Leslie 2005). I'll come to the experiments that provide strongest support for this view later (section 5.2 on page 12). Suppose we ignore challenges to such views and assume that something like this is correct.[8] Still, it would be a stretch to suppose that, in addition to tracking beliefs and desires with relatively simple contents (primarily concerning the locations of things), these abilities also allow one to track others' plans and subplans. More importantly, the representations of beliefs and other mental states discovered by this research (assuming for now that they have indeed been discovered) have limited effects on thought or action: these representations certainly influence eye movements (Clements & Perner 1994) and generate non-rational interference with judgements (Kovács, Téglás & Endress 2010); they may also shape some communicative actions (Liszkowski, Carpenter & Tomasello 2008) and to some extent guide word-learning (Carpenter, Call & Tomasello 2002) as well as modulating how subjects help others (Buttelmann, Carpenter & Tomasello 2009); but (as far as we know)

& Sommerville 2000). It may be that the understanding of goal-directed activity examined by these studies falls short of an understanding of intention.

[7] It is also worth noting that, on the whole, using dispositional measures of false belief understanding (e.g. asking children to deceive a character) rather than asking children to articulate judgements has no effect on the basic developmental picture (Wellman, Cross & Watson 2001; Polak & Harris 1999). It is reasonable to expect the same will hold for intention: the difficulty involved in tracking what others intend and believe is not primarily a difficulty with articulating judgements about mental states. (I come back to this from a different angle in section 5.2 on page 12.)

[8] In my view there are significant challenges to any such view: see Apperly & Butterfill (2009) and Butterfill & Apperly (2011).

they differ from full-blown representations of mental states in not playing the sort of role that knowledge or belief is thought to play in practical or theoretical reasoning.

What about Bratman's second point, [2]? Is this a good reply to either line of objection? The suggestion, [2], calls for phenomena which enable coordination of planning between agents but do entail the same conceptual and cognitive demands (whatever these in fact are) associated with tracking other's intentions. It is surely possible that, in a limited range of circumstances, agents might achieve coordination of planning not by representing each others' plans but instead thanks to a background of shared preferences, habits and conventions. But this will not be possible where the agents' aims are sufficiently novel, when the circumstances are unusual or when the agents are unfamiliar to each other. As far as I know, there is just one detailed account of how agents might have general abilities to track other's intentions—and that is by representing them. Given that we are thinking of intentions as elements of plans, this is bound to require conceptual sophistication (I'll come back to in what ways it might be cognitively demanding below). After all, if representing others's plans, the very things which tie their actions together over time, doesn't require conceptual sophistication, what does?

The first line of objection (conceptual sophistication) could be strengthened. Bratman already allows that there could be agents whose actions are purposive although they are not planning agents (p. 30: dogs and cats). Let's suppose that this kind of purposive action involves something intermediate between behaviour which merely has a function (perhaps exemplified by the behaviours of some insects) and full-blown intentional action as conceived by the planning theory. On the one hand, this kind of purposive action can involve novel goals (novel to the agents) and so cannot be straightforwardly explained by its evolutionary history. On the other hand no planning is involved. Could agents of this kind do things together such as go for a walk or lift a heavy sofa? There seems to be no reason to deny that they could, at least in principle. After all, providing that their success does not depend on coordinated planning or, alternatively, provided that their plans *in fact* coordinate—something which, in a limited but useful range of cases, could be taken care of by features of their environment and basic facts about their cognition (Richardson, Marsh & Baron 2007)—they will be able to do things together which they could not do alone. And we already have at least one account (mine) of this non-planning form of shared agency designed to complement Bratman's account of shared intention. So on the issue of conceptual sophistication, it seems to me that while there is not (yet) decisive empirical evidence on which to build an objection, and while planning agency seems not (yet) to have received much attention from cognitive and developmental psychologists, it is plausible to suppose that not all modest sociality involves dispositions to track others' plans.

There are some potential benefits to allowing for the possibility of non-planning shared agency. For one thing, abilities to have and act on shared intentions might be relevant to explaining *why* sophisticated forms of theory of mind cognition have emerged but they cannot be involved in explanations of *how* such cognition emerged since they presuppose it.[9] The existence of non-planning modest sociality means that forms of modest sociality could be relevant to explaining how sophisticated forms of theory of mind cognition emerge, and even how planning itself emerges. In general, only if there is non-planning shared agency can we suppose that modest sociality plays a really fundamental role in explaining how human cognition emerges in evolution or development (or ideally both).

So far I have distinguished two lines of objection, one concerning conceptual sophistication and the other concerning cognitive demands. I have suggested that Bratman's points [1] and [2] are not adequate replies to either objection, and that the first line of objection (conceptual sophistication) is promising but not decisive. In the rest of this section I want to focus just on the second line of objection, cognitive demands.

5.1. A crude objection

Philosophers (in conversation) are sometimes sceptical that ordinary humans know about others' knowledge of their intentions concerning others' intentions; Bratman's model over-intellectualises shared agency just as Grice's model of meaning, if taken as psychologically realistic, over-intellectualises communication. I think there is no justification for such brute scepticism and no commonsense or narrowly philosophical reason to think that shared agency is not both cognitively demanding and a pervasive feature of human life. Chapter 4 of Geurts (2011)[10] is a careful defence of the related claim that a Gricean theory of implicature is not *prima facie* psychologically implausible; in my view the same points apply to Bratman's model. (Geurts also argues that there is evidence in favour of a Gricean theory, but this part of his argument doesn't directly apply to Bratman's model of shared intention.)

5.2. Is shared agency cognitively demanding?

Would meeting the conditions involved in Bratman's model of shared intention be cognitively demanding in the ways specified above? I take it

[9] This claim needs qualifying; see Butterfill (2011a).

[10] This is the book I mentioned when we talked in the Study bar on Friday night.

that a positive answer by itself would not be an objection. But before considering a specific objection, it is worth asking whether is any evidence in favour of the hypothesis (if there were not we could stop here).

Although some researchers have strong theoretical commitments, so far there has been relatively little research on this topic. What there is generally but not universally supports the hypothesis (e.g. McKinnon & Moscovitch 2007; Apperly, Back, Samson & France 2008).[11] There is also some research suggesting that, in some real-time activities, adults are not invariably able to use ascriptions of mental states to interpret others' actions (Keysar, Lin & Barr 2003; Apperly, Carroll, Samson, Humphreys, Qureshi & Moffitt 2010). Finally, the hypothesis is indirectly supported by related developmental evidence, where a link between ascribing mental states (but not necessarily articulating such ascriptions) and executive function is well established (Perner & Lang 1999). On balance I would tentatively accept the hypothesis.

Just here a complication arises. Some recent studies with adults and infants suggest that, in some cases, tracking mental states including belief is unlikely to be cognitively demanding in the ways specified above. Infants, who have little working memory and limited executive function, are able to track other's beliefs in some situations (Onishi & Baillargeon 2005; Baillargeon et al. 2010; Southgate, Senju & Csibra 2007); and in adults, irrelevant facts about others' perceptions and beliefs appear to affect subjects' judgements about the number or location of an object (Samson, Apperly, Braithwaite & Andrews 2010; Kovács et al. 2010). In my view it is possible that these early-developing and possibly automatic abilities to track mental states are underpinned by processes which gain efficiency by representing not mental states as such but rather simpler, relational proxies for mental states (Apperly & Butterfill 2009; Butterfill & Apperly 2011; Surtees, Butterfill & Apperly 2011). If this is right (which I take to be an open question), abilities to track others' mental states without cognitive effort are likely be limited in ways that rule out tracking nested mental states (such as knowledge of intentions concerning intentions).

In short, then, there is no decisive reason to reject the hypothesis and some evidence in favour of it.

[11] Ferguson & Breheny (2011, p. 193) claim to show that 'complex higher-order ToM [theory of mind] inferences can be made without any greater discernable demands on costly cognitive processes [than comparable inferences not involving theory of mind]'. I think there are some objections to their conclusion (they are not always careful about the distinction between spontaneous and automatic processing, nor about the distinction between rapid and effortful processing), but there are also objections to research supporting the opposite conclusions.

5.3. An objection

Suppose that the hypothesis is true (i.e. meeting Bratman's sufficient conditions for shared intention is cognitively demanding in the ways identified above). Does it support an objection to Bratman? One such objection is made by Knoblich & Sebanz (2008, p. 2022), where the context (and personal communication) makes it clear their primary target it Bratman:

> '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.'

I note that this objection goes beyond the hypothesis. It would be consistent with the hypothesis to suppose that humans are able to track others' knowledge of their intentions concerning others' intentions in many real-time social interactions. After all, that a process places heavy demands executive function and working memory does not entail that it can't occur spontaneously in the space of a few seconds (as Geurts argues).

However, this objection can be strengthened a little. Suppose that Bratman's model provides universal coverage, i.e. that its conditions could be met in every case of shared agency. (The assumption of universal coverage would give us the strongest argument for the continuity thesis.) It follows that where the cognitive resources required for meeting Bratman's sufficient conditions for shared agency, whatever they are, are not available, there could not be shared agency. For instance, it might turn out to be a consequence of Bratman's view that agents counting backwards are typically unable to have and act on novel shared intentions. (Because the shared intentions are novel, pre-packaged coordination is ruled out.) Even if things did turn out this way, no objection follows immediately. There would only be an objection if we also knew that agents counting backwards (say) could engage in shared intentional activities.

I think a good way to avoid the vulnerability would be to admit the existence of non-planning shared agency (see section 2 on page 3). Or, to be more conservative, Bratman could allow that, if the evidence turns out to support this line of objection, then it shows that there is non-planning shared agency rather than that Bratman's model is incorrect. Invoking non-planning shared agency is helpful because, at least on some models of it (e.g. mine), the states involved do not impose the cognitive costs which (on Bratman's model) are associated with shared intention. Roughly, this is because non-planning shared agency rests on interlocking expectations concerning outcomes to which others' actions are directed rather than knowledge of interlocking intentions. So the existence of non-planning shared agency would make it more plausible that shared intention makes significant cognitive demands.

6. 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 2 on page 3).

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?

References

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: Evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*, 1093–1108.

Apperly, I. A. & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *2009*(116), 4.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*, 1201–1217.

Astington, J. (1991). Intention in the child's theory of mind. In D. Frye & C. Moore (Eds.), *Children's Theories of Mind: mental states and social understanding* (pp. 157–172). Hove: Erlbaum.

Astington, J. & Gopnik, A. (1991). Developing understanding of desire and intention. In A. Whiten (Ed.), *Natural Theories of the Mind: evolution, development and simulation of everyday mindreading* (pp. 39–50). Oxford: Blackwell.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342.

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

Butterfill, S. (submitted 2011b). What is joint action? a modestly deflationary account. `http://butterfill.com/what_is_joint_action/`.

Butterfill, S. & Apperly, I. A. (2011). How to construct a minimal theory of mind (submitted). `http://butterfill.com/papers/minimal_theory_of_mind/`.

Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology*, *20*, 393–420.

Clements, W. & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*(2), 705–717.

Ferguson, H. J. & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behaviour based on complex, conflicting desires. *Cognition*, *119*(2), 179–196.

Gergely, G., Nadasky, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.

Geurts, B. (2011). *Quantity Implicatures*. Cambridge University Press.

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

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

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.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830 –1834.

Leslie, A. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*(10), 459–62.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739.

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

McKinnon, M. C. & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218.

Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(8), 255–258.

Perner, J. & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*(9), 337–344.

Polak, A. & Harris, P. (1999). Deception by young children following noncompliance. *Developmental Psychology*, *35*(2), 561–568.

Richardson, M. J., Marsh, K. L., & Baron, R. M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance. Vol. 33(4)*, *33*(4), 845–859.

Samson, D., Apperly, I. A., Braithwaite, J. J., & Andrews, B. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255–1266.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science*, *18*(7), 587–592.

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2011). Direct and indirect measures of level⊠2 perspective⊠taking in children and adults. *British Journal of Developmental Psychology*.

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.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development*, *72*(3), 655–684.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.

Woodward, A. L. & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, *11*(1), 73–77.