

Shared Agency Involves Changing Perspective: A Counterexample to Bratman

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

December 8, 2012

Abstract

The leading, best developed account of shared agency, Michael Bratman's, hinges on the claim that, roughly, we have a shared intention if we each intend that we J and also that we J by way of these intentions and meshing subplans of them where this is all common knowledge. This paper provides a counterexample to the sufficiency of this condition for shared intention. The counterexample arises because it is possible for agents to meet the condition, and to act rationally on the specified intentions and knowledge, while conceiving of each other's actions only as constraints to work around and opportunities to exploit. Shared agency requires more than this, or so we argue. We also suggest a way of revising Bratman's account that is consistent with his general approach. In some or all cases, shared agency differs from individual agency in part because shared agency involves changing perspective so that one's own and other's actions are conceived of as parts of a single plan.

1. Shared Intention

Why, if at all, is a notion of shared intention needed? This question is standardly answered by appeal to contrast cases (compare Bratman 2009, p. 150). Thus Gilbert (1990) contrasts friends intentionally walking together with two people who happen to be walking side by side. And Searle (1990) contrasts park visitors who simultaneously run to a central shelter in performing a dance with park visitors who likewise run to the central shelter but only because of an impending storm. These and other contrast cases invite the question, How do cases involving shared agency differ from cases involving only parallel individual agency?

The first contrast case, Gilbert's, shows that the difference can't be just a matter of coordination because people who merely happen to be walking side by side each other also need to coordinate their actions in order to avoid colliding. Note also that in both cases each individual's walking is intentional, so our intentionally walking together cannot be only a matter of our each intentionally walking. The second contrast case, Searle's, shows that the difference can't just be that the resulting actions have a common effect because merely parallel actions can have common effects too.¹ Perhaps, then, a notion of shared intention is needed to distinguish the two cases. Perhaps it is our acting on a shared intention that we walk together which distinguishes us from two strangers who happen to be walking side by side.²

But what could shared intention be? In an influential series of papers,³ Bratman claims that the following are collectively sufficient⁴ conditions for you and I to have a shared intention that we J:

- '1. (a) I intend that we J and (b) you intend that we J
- '2. I intend that we J in accordance with and because of la, lb, and meshing subplans of la and lb; you intend that we J in accordance with and because of la, lb, and meshing subplans of la and lb
- '3. 1 and 2 are common knowledge between us' (Bratman 1993, p. View 4).

In this paper we give a counterexample to Bratman's claim that the above conditions, (1)–(3), are collectively sufficient conditions for shared intention. We shall also suggest a revision to avoid the counterexample.

Before going further we must distinguish two versions of the claim that (1)–(3) are collectively sufficient for shared intention. The *weak claim* is that there is some J such that these conditions are sufficient for you and I to intend that we J. The *strong claim* is that for any J, these conditions are sufficient for you and I to intend that we J. Our initial counterexample will be directed to the strong claim. However, after we have constructed this counterexample it will become clear that there are also counterexamples involving a wide

¹ This use of contrast cases resembles Pears (1971): he uses contrast cases to argue that whether something is an ordinary, individual action depends on its antecedents.

² Many philosophers hold that a notion of shared intention is useful for understanding acting together. Compare Gilbert (2006, p. 5): 'I take a collective action to involve a collective intention.' See also Carpenter (2009, p. 381), Call (2009, p. 369), Kutz (2000), Rakoczy (2006, p. 117) and Tollefsen (2005).

³ See Bratman (1992, 1993, 1997, 2009). For influences beyond philosophy, see e.g. Tomasello et al. (2005) and Knoblich & Sebanz (2008).

⁴ In Bratman (1992), the following were offered as jointly sufficient *and individually necessary* conditions; the retreat to sufficient conditions occurs in Bratman (1997, pp. 143–4) where he notes that 'for all that I have said, shared intention might be multiply realizable.'

range mundane activities to which Bratman explicitly takes his account to apply. Even walking together⁵ and painting a house together⁶ can serve as counterexamples.

To show that Bratman's conditions, (1)–(3), are not in fact sufficient for shared intention we need a case where the conditions are met and we lack a shared intention. But how could we determine that we lack a shared intention? As already mentioned, the notion of shared intention is supposed to make it possible to characterise systematically a difference between cases involving shared agency (such as our walking together) and cases involving only parallel individual agency (such as two strangers who happen to be walking the same route side-by-side). Suppose, then, that we had a trio of cases, A, B and C, each involving two agents. Suppose, further, that A and B involved parallel agency only, whereas C involved shared agency. Then we could be sure that A and B do not involve shared intention. Now suppose that, for some J, Bratman's conditions, (1)–(3) above, were met in cases B and C alike. Suppose also that in each case, B and C, the structure of intention and knowledge mention specified by Bratman's conditions played an appropriate role in guiding the agents' actions, and that in each case the agents did thereby successfully J. Finally, suppose that cases A, B and C are as similar as possible except for the differences mentioned. Then case B would be our counterexample: Bratman's conditions are met but there is no shared intention. This is how our counterexample will work. (Strictly speaking, case A is not necessary; but including it simplifies exposition.)

Several preliminaries are necessary for the construction of our counterexample. These preliminaries might easily give the impression that our counterexample depends on an artificial settings. However, having introduced the primary counterexample in an artificial setting, we will go on to show that counterexamples can also be constructed for mundane activities including walking together.

2. Neutral with Respect to Shared Intentionality

Before we can introduce the counterexample we need to highlight a feature of Bratman's account which we shall be exploiting.

Consider the contents of the intentions concerning our J-ing in the above clauses, (1)–(3). What sort of activity can you intend when you intend that we J? We cannot restrict possible values of J to activities which involve shared agency. In imposing any such restriction we would be assuming the very notion that an account of shared intention is supposed to illuminate.

⁵ See Bratman (2009, p. 150).

⁶ See Bratman (1992, p. 331).

Rather the above conditions, (1)–(3), must be sufficient for shared intention even for some values of J which are ‘neutral with respect to shared intentionality’.⁷

To illustrate, first consider Bratman’s ‘mafia case’ where two people go to New York together by virtue of one of them forcing the other into a car and driving off (Bratman 1992, p. 333). Next consider a paradigm case of where two friends cooperatively go to New York together. It’s possible to conceive of an act type of going to New York together which is broad enough that both cases, the coercive and the cooperative, fall under it. And it’s even possible to desire or intend to be involved in an act of this type. A gangster might intend to go to New York with someone without having yet decided whether they will do this cooperatively or coercively.

A consequence is that, in the right situations, one of us can rationally intend that we J , and can intend this unilaterally, that is without depending on anyone else intending that we J . Suppose you know that I am going to New York via a certain route at a particular time, and that I will do this regardless of what you do. Suppose also that you can rationally intend that you go to New York in the same manner, and that you know that if you act on this intention the upshot will be that we will go to New York together (although I may not know that we are going together—perhaps you will conceal your presence from me). Then you can rationally intend that we go to New York together, and you can intend this irrespective of whether I have any corresponding intention—providing, of course, that in so intending you are conceiving of our going to New York together in a way that is neutral with respect to shared intentionality. What follows depends on the premise that this is indeed possible.⁸

3. Unshared Intentions

As a further preliminary we need to introduce a definition. Let us stipulate that we have an *unshared intention* that we $\langle J_1, J_2 \rangle$ where $J_1 \neq J_2$ just if:

- 1'. (a) I intend that we J_1 and (b) you intend that we J_2
- 2'. I intend that we J_1 in accordance with and because of la, lb, and meshing subplans of la and lb; you intend that we J_2 in

⁷ Bratman (1997, p. 147). This refines Bratman’s earlier view that some admissible values of J are cooperatively neutral where an act-type is *cooperatively neutral* just if ‘joint performance of an act of that type may be cooperative, but it need not be’ (Bratman 1992, p. 330).

⁸ Bratman (1997) defends this claim at length. Note also that this claim must be true if Bratman’s account of shared intention is to provide an informative and systematic distinction between the contrast cases mentioned at the start.

accordance with and because of la, lb, and meshing subplans of la and lb

3'. 1 and 2 are common knowledge between us.

In defining unshared intention we have used conditions exactly like Bratman's sufficient conditions for shared intention except that Bratman's conditions have $J_1=J_2$. At this point it might be natural for readers to suppose that agents could not have unshared intentions, or at least that they could not do so without irrationality. In this section we describe a possible situation in which two agents have an unshared intention without irrationality, deception or even ignorance. This possible situation is not the promised counterexample, but it does form the basis for it.

Let us first introduce the activity we shall focus on. Ayesha and Ahmed are playing a simple video game which involves moving a cross around a two-dimensional space littered with barriers. Ayesha can only accelerate the cross backwards or forwards, while Ahmed can only accelerate it left or right. The cross moves around and interacts with the barriers in ways both players can predict. The players are given tasks independently. These tasks always involve making the cross hit a target within two minutes of starting. A player succeeds when the cross hits her target, regardless of what happens to the cross afterwards. (It may go on to hit another target.) In this case, Ayesha's task is to make the cross hit the red square while Ahmed's task is make the cross hit the blue circle. In general it is possible that either or both will succeed, or that they will both fail. Each movement carries a small cost to the player who moves, so that Ayesha and Ahmed each attempt to minimize how much he or she moves the cross consistently with completing his or her task. At the outset, Ayesha and Ahmed are each neutral on whether the other succeeds or fails. They are not opponents and do not seek to undermine each other's efforts, but each is entirely concerned with his or her own task. All of this is common knowledge for Ayesha and Ahmed. They both know who has which task, what constraints they face and what their motives are.

Consider the possibility of one player intending, unilaterally, that the two players do something. Suppose that one of the players—Ayesha, say—can knowledgeably predict that if she performs a certain sequence of actions, $\langle a_1, a_2, \dots a_n \rangle$, then Ahmed will simultaneously perform certain other actions, $\langle b_1, b_2, \dots b_n \rangle$, and the upshot will be that the cross hits the red square. Were this to happen, it would be true that Ayesha and Ahmed made the cross hit the red square. Suppose, further, that Ayesha can intend to perform those actions $\langle a_1, a_2, \dots a_n \rangle$. Then Ayesha can intend, unilaterally, that they, Ayesha and Ahmed, make the cross hit the red square. (This intention concerns an action conceived of as neutral with respect to shared intentionality; see Section 2 on page 3).

Unshared intentions require a kind of symmetry. Let us suppose that the above sequences of actions, Ayesha's $\langle a_1, a_2, \dots a_n \rangle$ and Ahmed's $\langle b_1, b_2, \dots b_n \rangle$, will also result in the cross hitting the blue circle. (Since the cross has momentum, we can suppose that it will hit both the red square and the blue circle at some time after these action sequences have been performed.) Then by the reasoning just offered, Ahmed could intend that they, Ayesha and Ahmed make the cross hit the blue circle. So Ayesha and Ahmed could meet the first condition, (1), for having an unshared intention.

What about the second condition, (2)? Suppose that Ayesha knows two further things. First, that Ahmed intends that they, Ayesha and Ahmed, make the cross hit the blue circle. Second, that in acting on his intention Ahmed will perform actions $\langle b_1, b_2, \dots b_n \rangle$. Then Ayesha can intend that they, Ayesha and Ahmed, make the cross hit the red square in accordance with and because of her intention that they make the cross hit the red square and in accordance with and because of Ahmed's intention that they make the cross hit the blue circle.

This is not quite enough to meet the second condition, (2), because there is also a requirement about meshing subplans. To apply this requirement we need to generalise Bratman's definition of meshing:

‘our individual subplans concerning our J-ing *mesh* just in case there is some way we could J that would not violate either of our subplans but would, rather, involve the successful execution of those subplans’ (Bratman 1993, p. 106).

A natural generalisation is this:

our individual subplans concerning our $\langle J_1, J_2 \rangle$ -ing *mesh* just in case there is some way I could J_1 and you could J_2 that would not violate either of our subplans but would, rather, involve the successful execution of those subplans.

To illustrate, Ayesha's and Ahmed's subplans would fail to mesh if, in intending that they make the cross hit the red square, Ayesha's plans had included pushing Ahmed out of the way and seizing his controls. They would also fail to mesh if Ayesha were planning to trick Ahmed into a situation where he would be unable to perform the actions he had been planning. But in the case we have been describing, the agents' subplans mesh perfectly. Each agent's subplans involve manipulating his or her own controls, and the successes each seeks in doing this depends on the other successfully carrying out their subplans. So Ayesha can rationally intend that they, Ayesha and Ahmed, make the cross hit the red square in accordance with and because of their intentions and meshing subplans of them. And Ahmed likewise for making the cross hit the blue circle.

The only outstanding requirement for Ayesha and Ahmed to have an unshared intention is that their various intentions are common knowledge. Assuming common knowledge is possible concerning Bratman's conditions (1)–(2) on page 2, it is likewise possible for the corresponding conditions on unshared intention, (1')–(2') on page 4. So Ayesha and Ahmed can have an unshared intention that they $\langle J_1, J_2 \rangle$ where J_1 is Ayesha and Ahmed's making the cross hit the red square and J_2 is their making the cross hit the blue circle.

So far we have shown that it is possible for two agents to have an unshared intention without irrationality, deception or ignorance. Of course unshared intentions may be rare. But what matters for our counterexample is just that they are possible.

4. The Counterexample

In the situation just described, Ayesha and Ahmed are playing a game and have different tasks. Thanks to special features of the game environment, they both succeed by acting on an unshared intention. Now compare two further players, Beatrice and Baldric, who are playing the same game. Their situations, knowledge states, intentions and actions are as similar as possible to Ayesha's and Ahmed's except for one detail. Just by chance they have been assigned identical tasks: Beatrice's task is to make the cross hit the red square and Baldric's task is the same. So where Ayesha and Ahmed have an unshared intention that they $\langle J_1, J_2 \rangle$, Beatrice and Baldric meet Bratman's conditions (1)–(3) for having a shared intention that they J_1 . But Beatrice, in planning and acting, does not rely on the coincidence of their intentions; and nor does Baldric. (Beatrice relies on the fact Baldric intends that they J_1 , of course; but she does not rely on the fact that Baldric intends what she intends.) Let us further stipulate that, due to an artefact of the way the game is structured, the unshared intention and the Bratman intention (as we might label the structure of intention and knowledge while leaving open whether it constitutes a shared intention) result in the two pairs performing same actions in the same way. That is, Beatrice reasons about Baldric much as Ayesha reasons about Ahmed and Beatrice and does what Ayesha does; and likewise for Baldric and Ahmed.

We claim that Beatrice and Baldric have a shared intention that they J_1 only if Ayesha and Ahmed have a shared intention.⁹ This claim follows from

⁹ Strictly speaking, what matters for our argument is whether Ayesha and Ahmed have a shared intention *in virtue of having the unshared intention that they* $\langle J_1, J_2 \rangle$. This is because, strictly speaking, we need to show, not that Beatrice and Baldric lack any shared intention whatsoever, but only that they lack a shared intention that they J_1 just in virtue of meeting Bratman's conditions (see (1)–(3) on page 2). For ease of exposition this is not

the similarities of the two cases. The only difference is that Beatrice and Baldric happen to be assigned the same task, whereas Ayesha and Ahmed are not. And neither Beatrice nor Baldric makes use of the fact that they have the same task. (This is not due to ignorance: it's just how they choose to approach their tasks.) So if we consider how Beatrice and Baldric's case differs from that of Ayesha and Ahmed, we can see that these differences do not plausibly amount to a difference with respect to shared agency. Shared intention cannot feature in one case but not the other.

To show that Beatrice and Baldric do not have a shared intention it remains only to show that Ayesha and Ahmed do not have one. Here we must be careful. First note that, since Bratman provides only sufficient conditions for shared intention, his account doesn't tell us that an unshared intention is not a shared intention. For all Bratman says, the conditions defining unshared intention might be sufficient for shared intention.

But doesn't shared intention require at least this much, that there be a single activity about which each the agent involved has an intention? This might reasonably be doubted by those who, like Bratman, reject the Simple View according to which when an individual intentionally J-s she has an intention concerning her J-ing (Bratman 1984). So we shall not infer that Ayesha and Ahmed lack a shared intention just because (by construction) there is no J such that Ayesha and Ahmed each intend that they, Ayesha and Ahmed, J.

Can we then appeal directly to intuition to show that Ayesha and Ahmed lack a shared intention? Ayesha sees Ahmed's actions as constraints on her own, or else as opportunities. She exploits Ahmed's intentions for her own ends. Of course the situation is reciprocal: Ahmed exploits Ayesha in equal measure. Each allows himself or herself to be exploited by the other because being exploited enables exploiting, and this is the full extent of their cooperation. We don't suppose that reciprocal exploitation is incompatible with shared intention. But Ayesha and Ahmed's interaction consists entirely in this sort reciprocal exploitation, where each agent sees the other's actions only as constraints or opportunities. Perhaps it is clear enough that the sort of shared agency that an account of shared intention is supposed to capture must involve more than this. If so, we can already claim that Ayesha and Ahmed are not acting on a shared intention. But philosophers' intuitions about shared agency may vary, so it would be better if we could avoid such a blunt appeal to intuition.

How else could we support the claim that Beatrice and Baldric lack a shared intention? As mentioned at the start, the contrast cases are often used to anchor intuitions in theorising about shared agency. In the next section we shall further support our claim by contrasting Beatrice and Baldric's

made explicit in the main text.

case with a further case, one which is as similar as possible and which does seem to involve shared agency. The fact that Beatrice and Baldric's case contrasts with this new case will support the claim that Beatrice and Baldric lack shared intention. This is the aim of the following section. A further motive for introducing this new contrast case is that it provides an indication concerning what is missing from Bratman's account, and so may help us to understand why his conditions are not sufficient for shared intention.

5. Individually planning all of the agents' actions

Above we introduced what will turn out to be a counterexample to Bratman's account, the case of Beatrice and Baldric. But we have yet to show that this case really is a counterexample. For all we have said so far, a proponent of Bratman's view might insist that Beatrice and Baldric do have a shared intention that they J_1 . To show that they do not, we shall contrast Beatrice and Baldric's case with a third case that is as similar as possible but does involve shared intention.

First we need to introduce the notion of an agent-neutral plan and some observations. We stipulate that a planning process, or a plan, is *agent-neutral* just if it does not involve identifying any particular agents. This sort of planning is quite common. For example, some housemates who have decided to make a pizza might sit down together to plan what needs doing without yet assigning roles to particular individuals. Someone needs to prepare the dough, another person should prepare toppings, while a third might mix some salad. In so planning, each housemate is thinking about what is to be done and not what she herself will do. At some point the housemates stop planning. (This does not necessarily mean that they have a fully worked out plan; like any other plans, agent-neutral plans can have gaps that may need filling in later.) They now divide up the roles. Of course they may not find a way of dividing up roles that everyone is prepared to go along with—individuals' preferences, abilities and intentions may block the plan's adoption. But suppose the housemates do divide up the roles in a way that is acceptable to everyone, and that each implements her part in the plan. Then each conceives of her own and the others' actions as part of single plan directed to achieving a single outcome.

In this example the housemates plan together and agree on a common plan. Planning together is plausibly an activity which involves shared agency. Note, however, that an individual can construct an agent-neutral plan by herself, even if its eventual execution will involve others. In fact, two or more individuals who are assigned the same task and who will eventually collaborate might first each individually engage in agent-neutral planning and then individually assign roles. And they might do this in isolation from

each other, without discussion or communication.

Where there is an outcome and some agents each individually, without discussion or communication, plan for the outcome and where each agent has a single plan specifying roles for herself and all of the other agents, we shall say that these agents each *individually plan all of the agents' actions*.¹⁰ Agents can do this by first constructing agent-neutral plans and then assigning roles to themselves and the other agents. So, to return to our earlier example, if the housemates, having decided to make a pizza, had each individually devised agent-neutral plans for the construction of the pizza and then each individually assigned roles in their own plans to themselves and the others, then they would have each individually planned all of the agents' actions.

For all we have said, agent-neutral plans might not be involved in all cases where some agents each individually plan all of the agents' actions. We invoke agent-neutral plans only as an aid to showing that the notion of some agents each individually planning all of the agents' actions is coherent.

If some agents each individually plan all of the agents' actions, could they nevertheless know that they will form plans that are sufficiently similar or jointly viable?¹¹ Consider housemates making a pizza together, where everyone knows roughly the same recipe and differences are unlikely to matter. Or consider two agents moving a table from one room to another together, where there is an obviously optimal route and the agents' positions manifestly determine which ends of the tables they should hold. For these outcomes and situations, the agents' expertise and planning strategies may mean that each could knowledgeably predict that they will form sufficiently similar agent-neutral plans. And sometimes commonly known properties of the agents, such as their distribution in space or their sizes, mean that there is exactly one most obvious way of dividing up roles. In any such case,

¹⁰ This phrase is ambiguous (in a way we know no way concise of avoiding). The idea is that each agent has a single plan in which she and all other agents are assigned roles. There is no requirement that any agent's plan must specify all the actions which the agents between them will perform.

¹¹ Suppose that, for some outcome, two or more agents each have a plan for the realisation of that outcome. We stipulate that these plans are *jointly viable* just if the plan which involves each agent performing the roles she is assigned in her own plan is acceptable to all the agents and a viable means of realising the outcome. For an example of jointly viable but significantly different plans, suppose that our task is to transport some apples and pears to a warehouse. My plan involves you taking care of the apples while I put the pears into the yellow crates and load them into the van through the side door. If your plan involves me taking care of the pears while you put the apples into the green crates and load them into the van through the rear doors, then our plans are compatible. Consider also a case in which our task is to press a button simultaneously. If your plan is that each of us will start to act in exactly 60 seconds whereas my plan specifies that you are the leader and we will start to act when you do, then our plans are jointly viable.

the agents who each individually plan all of the agents' actions may form sufficiently similar plans and know that they will do so.

Agents who will collaborate in some way, such as by making a pizza together, can coordinate their actions by each individually planning all of the agents' actions. Suppose some agents are about to act together in order to bring about some outcome. They know (we stipulate) that if they each individually plan all of the agents' actions, they will form sufficiently similar plans. Each agent also knows enough about herself and the others to be able to determine, without communicating, that the resulting plans will be acceptable to everyone. And all of this is common knowledge to the agents. Then these agents might individually plan all of the agents' actions, and each might rationally perform her part in her plan, knowing that the others will do likewise. In this way, some agents' individually planning all of the agents' actions can rationally coordinate their plans and actions.¹²

But why would some agents each individually plan all of the agents' actions? After all, no agent gets to perform others' actions (or, if they do, this is exceptional). So why would anyone individually plan others' actions, knowing that she won't perform them? In many cases, a plan for one agent's actions would need to be constrained by, and to constrain, plans for other agents' actions. When, for example, two people are about to move a heavy table from one room to the next, two separate plans for their individual actions would have to involve compatible timings, forces and routes. In principle, each table-lifter might form an initial plan for her own actions only. Each could use her initial plan to make predictions about the other's likely actions. She could then revise the plan for her own actions in the light of these predictions. This cycle of prediction and revision might be repeated. The plans might also involve conditional elements to accommodate unpredicted actions. So when an agent plans only her own actions, she will sometimes need to rely on predictions about what others will do, and these predictions will sometimes depend on what the agent plans to do. Another way to proceed would be for the agent to plan all of the agents' actions, perhaps by first making an agent-neutral plan and then assigning roles. So the reason why some agents might each plan all of the agents' actions is that doing so provides one way to solve a coordination problem. We don't claim that this is the only or best way to coordinate. Our claim is just that some agents sometimes achieve coordination in part by means of each agent's individually planning all of the agents' actions.¹³

¹² Note that this coordination is achieved without presupposing that the agents exercise shared agency. Of course, if some agents individually plan all of the agents' actions and then each rationally performs her part in her own plan, the upshot might be an exercise of shared agency. But this shared agency is a consequence of, not a requirement on, their individually planning all of the agents' actions.

¹³ As an aside, we note that Tsai et al. (2011) provide evidence indicating that planning for

Now recall Beatrice and Baldric (from Section 4 on page 7). We said earlier that they each see the other's actions as constraints to work around or opportunities to exploit only. This is because each agent plans only her own actions. Neither agent plans for both of their actions and so neither conceives of both of their actions as parts of a single plan. In constructing a case that contrasts with Beatrice and Baldric's we shall rely on the distinction between a plan that includes both one's own and others' actions and a plan that merely exploits, and is constrained by, predictions about the others' actions. This distinction matters independently of whether two such plans would result in different actions. It matters because forming a plan for your own and others' actions involves changing perspective. That is, it involves conceiving of the others' actions not only as constraints and opportunities, but as on a par with your own actions insofar as your actions and theirs are elements in a single plan directed to a single outcome.

In the case of ordinary, individual agency, there is a familiar and important distinction between merely predicting one's own actions and planning them.¹⁴ It is possible and perhaps sometimes useful to make predictions about one's future actions, of course. But to rely on predictions about one's own actions in planning can involve a kind of estrangement from oneself (compare Moran 2001). We have just been arguing that in the case of shared agency, the distinction between predicting and planning actions also applies to the actions of agents other than oneself.

Applying this distinction in cases of shared agency requires additional subtlety, of course. After all, one agent cannot rationally include another's actions in a plan she knows she will act on unless she could predict the other's actions. (Or unless she knows the other agent will act directly on the first's intentions—see Roth (2010).) This fact should not obscure the distinct ways in which one conceives of another agent's actions in planning and predicting them, however. For, as we have argued, in some cases planning another agent's actions does not require actually predicting them. In some cases, planning another's actions does not depend on having predicted them but rather provides a ground for prediction. This is why there is a difference in perspective between merely predicting another's actions and conceiving of them as on a par with one's own in being elements of a single plan.

Now that we have some background on how some agents might each individually plan all of the agents' actions, we can introduce a third and final case. This case needs to be as similar as possible to Beatrice and Baldric's while involving shared agency.

one's own and others' actions happens not only in conscious practical reasoning but also in planning that we may not ordinarily be directly aware of.

¹⁴ Even Velleman, who argues that '[d]eciding on an action ... is just a peculiar way of predicting it' (1985, p. 48), can accommodate a distinction between merely predicting and planning actions (see ?, p. 195).

Caitlin and Ciaran start in the same situation as Beatrice and Baldric. Each is tasked with making the cross hit the red square (J_1). Once again each cares only about her own success at the outset. Caitlin, knowing that she and Ciaran have the same task, takes the view that the best way for her to succeed is plan how two agents in their situation could J_1 . So Caitlin ends up with an agent-neutral plan. She next assigns one role in the plan to herself and the other to Ciaran. At this point Caitlin considers whether she would be prepared to go along with the plan given her intentions, preferences and values, and she also considers whether Ciaran would be prepared to go along with it too. In this case it happens that both would be prepared to go along with the plan. Caitlin then knowledgeably predicts that Ciaran, who has similar planning abilities and has been approaching their task in a similar way, will have made a sufficiently similar plan. Finally, Caitlin attempts to carry out her part in her plan; and Ciaran does likewise.

In short, then, Caitlin and Ciaran are like Beatrice and Baldric in nearly every respect. Each pair has a Bratman intention that they J_1 ,¹⁵ each pair acts on this intention and each pair ends up performing the same sequence of actions. The difference is just that Beatrice and Baldric make no use of the fact that they are performing the same task. This fact does not feature in their planning. Rather, each plans her own actions only and treats the other's actions as constraints to work around or opportunities to exploit. By contrast, Caitlin and Ciaran embrace the fact that they are performing the same task. They may not like having to act together; in fact each may far prefer to act alone were that possible. And, like Beatrice and Baldric, they are unconcerned with each other's success except insofar as their own success depends on it. But Caitlin and Ciaran nevertheless make use of the fact that they have to perform a single task together by each constructing a single plan including both of their actions and then carrying out their parts in these plans. So why does Caitlin and Ciaran's case, but not Beatrice and Baldric's, plausibly involve shared agency? The reason is this: at some stage of the planning which rationally guides and coordinates their actions, Caitlin and Ciaran each conceive of both of their actions as parts of a single plan directed to achieving a single outcome.

Given the way Beatrice and Baldric's case contrasts with Caitlin and Ciaran's, we conclude that Beatrice and Baldric do not have a shared intention. But, by construction, Beatrice and Baldric do meet Bratman's conditions for shared intention, and they do act appropriately on the corresponding intentions and knowledge. So Beatrice and Baldric's case is a counterexample to sufficiency of Bratman's conditions for shared intention.

In describing Caitlin and Ciaran we have not introduced contralateral

¹⁵ As stipulated above (see page 7), two agents have a *Bratman intention* that they J just if they meet conditions (1)–(3) on page 2.

commitments or any other element foreign to Bratman's account of shared agency. Our counterexample draws on the same planning resources Bratman's account draws on, with just one addition. The addition is the idea that an agent might plan her own actions only or else might plan both her own and others' actions. Agents often conceive of others' actions as external to their own plans, and so as either entirely irrelevant or constraints and opportunities only. But shared agency sometimes or always involves agents who each individually plan all of the agents' actions and so conceive of their own and the others' actions as parts of a single plan. In at least some cases, shared agency can be distinguished from individual agency in part because it involves changing perspective in this way.

6. Mundane Counterexamples

Beatrice and Baldric's case is a counterexample to the claim that for any J, Bratman's conditions, (1)–(3) on page 2, are collectively sufficient for us to have a shared intention that we J. (We called this the 'strong claim' on page 2.) If this counterexample were an isolated case, it might be tempting to suppose that it can provide little insight into shared agency. Could proponents of Bratman's view avoid counterexamples by claiming that the his conditions are sufficient for mundane exercises of shared agency such as those involved in walking together or painting a house together? In this section we show that they could not. Given that Beatrice and Baldric's case is a counterexample, there are also many further counterexamples involving mundane activities and less elaborate props.

Here is how Beatrice and Baldric walk together. They are firmly tied at the ankle, and neither is strong enough to move without the other. Beatrice needs to get to the corner, and so does Baldric. Further, they have a Bratman intention that they walk to the corner, and they act on this intention in walking to the corner. As before, each plans her own walking in the light of her predictions about how the other will walk, treating the other's actions as constraints and opportunities. This is not a case of shared agency. The reason is not that Beatrice and Baldric are tied together against their wills; after all, many cases of shared agency involve agents who are unwillingly bound to each other. The reason their walking does not involve shared agency is rather that each conceives of the others' intentions and actions only as constraints and opportunities. Beatrice has a plan for what she, Beatrice, will do to ensure that they arrive at the corner; and Baldric likewise plans what he only will do. Each agent's plan relies on predictions about how the other will act and contains subplans which are conditional on the other's subplans and actions. But this involves no more than the sort of interdependence that might hold between two plans directed at different ends, as in cases of un-

shared intention. Beatrice and Baldric fail to make any use of the fact that they each intend the same thing, namely that they walk to the corner.

For another counterexample, consider house painting. First imagine that Ayesha and Ahmed are tasked with painting the outsides of adjacent houses on the same day, which happens to be windy. Each wants their house to be painted a single colour, and would prefer that it not be painted at all rather than that it be painted in several colours. But because they are spraying the paint onto the houses in blustery winds, neither can avoid spraying a significant quantity of paint onto the other's house as well as their own house. In one respect this is fortunate. For they will each lose so much paint from spraying in windy weather that, normally, neither would have enough paint to cover her house. But if they each make use of wind-borne spray from the other, both can succeed. They therefore form an unshared intention concerning their house painting projects. (The notion of unshared intention is explained in Section 3 on page 4.) In acting on this unshared intention, Ayesha starts painting first, using a colour which she knows is acceptable and available to Ahmed. Ayesha and Ahmed then each plan a route around their own house which will keep them out of the path of the other painter while also saving paint by taking advantage of foreseeable wind-borne paint from the other.

Ayesha and Ahmed's house painting involves only unshared intention and is therefore not a counterexample. But now imagine that there is a single house which Beatrice and Baldric are each motivated to paint. Because neither has sufficient paint for the job, they need to work together if they are to succeed. They therefore form a Bratman intention that they paint the house. But in acting on this intention, they plan and act much as Ayesha and Ahmed do. Each agent plans what she will do to ensure that they paint the house, treating the other's intentions and actions as constraints and opportunities. Although the agents' plans and intentions interlock, the agents do not conceive of their actions as parts of a single plan. From either agent's point of view, their intentions and actions are no more intimately related than are Ayesha and Ahmed's. This is not shared agency.

Beatrice and Baldric's attitude may seem unnatural. Perhaps it is hard to believe that ordinary people in their situation would fail to exploit the fact that they have the same task. This may be why Bratman's conditions appear, misleadingly, to be sufficient for shared intention.

7. How to Fix Bratman's Account

How could Bratman's account be revised to avoid the counterexample? We might try strengthening Bratman's requirement about meshing subplans of intentions. This requirement is that, concerning our shared intention that

we J, we each intend that we J in accordance with our intentions that we J and meshing subplans of them (see (2) on page 2). The case of Beatrice and Baldrick suggests that this requirement is not strong enough. And the case of Caitlin and Ciaran suggests that we might try strengthening it by appealing in some way to the idea that each agent can have a plan specifying roles for all of the agents.

One way to strengthen Bratman's account would be by adding a fourth condition his three conditions. An alternative, which we prefer, is to introduce a new restriction on how agents who have a shared intention may conceive of J in intending that they J. As we have seen, Bratman's account requires only that, in some cases, the agents conceive of J in a way that is neutral with respect to shared intentionality (see Section 2 on page 3). We shall suggest a further requirement that is compatible with this one.

We stipulate that an activity is *distributed* just if:

- a) there is a single outcome, G, and two or more agents each have a plan for G which specifies a role for all of the agents;
- b) these plans are identical or compatible,¹⁶ and they are each acceptable to all;
- c) each agent performs the roles assigned to her in her own plan;
- d) the activity consists in nothing other than the agents performing these roles; and
- e) by performing these roles, the agents G.

In specifying the class of distributed activities we have not made direct appeal to shared agency. Nor is any covert appeal required, as our earlier discussion should make clear (see Section 5 on page 9). To use Bratman's term, distributed activities can be conceived of in ways that are neutral with respect to shared intentionality (see Section 2 on page 3). So we can appeal to the notion of a distributed activity in explicating shared intention without circularity.

Many ordinary interactions involve distributed activities. Two housemates have agreed to make a pizza. Each start with an idea of what needs doing. One starts by preparing the dough, so the other peels, washes and chops the vegetables. As the activity unfolds in this way, each gradually adds details to her plan and assigns roles.

Distributed activities can be simple. In walking to a metro station together, two friends may each form and act on a plan that specifies some details of both of their actions. This is enough for their activity to be distributed in the above sense.

¹⁶ The notion of compatibility is defined in footnote ?? on page ??.

When two or more agents each intend that they do something, they may conceive of the intended activity as a distributed activity. Where this happens, part of what the agents intend is, in effect, that they will each, at some stage of their planning, construe their own actions as part of a single plan which also involves the others' plans.

We can exploit this possibility to strengthen Bratman's conditions for shared intention. We leave the conditions for shared intention exactly as Bratman specifies them (see (1)–(3) on page 2) but add the further requirement that the agents must, in intending that they J, each conceive of J as a distributed activity.

Thus revised, the account provides conditions that are met in the case of Caitlin and Ciaran but not Beatrice and Baldrick, as required. We have not shown, of course, that this modification to Bratman's account will enable it to avoid any other counterexamples. Perhaps further requirements are necessary. Or perhaps it is impossible to give informative sufficient conditions for shared intention. Our aims here were only to show that Bratman's account faces counterexamples, and to suggest a way of revising it without abandoning core features of Bratman's approach to shared intention.

8. Conclusion

We started with some contrasts between shared agency and parallel but merely individual agency. These provide an intuitive fix on the notion of shared intention: shared intention, whatever it is, is what distinguishes thinks like two friends walking somewhere together from things like two strangers who walk the same route side by side. By introducing new contrast cases, we provided a counterexample to the view that a certain interlocking structure of intention and knowledge is sufficient for shared intention. Two agents might lack a shared intention even though they meet conditions Bratman offers as sufficient for shared intention. That is, they might lack a shared intention even though it is common knowledge to them that they each intend that they J and intend that they J in accordance with and because of these intentions and meshing subplans of them.

Bratman's conditions are not sufficient for shared intention because they leave open the possibility that two or more agents could meet his conditions while each failing to make use of the fact that there is a single outcome to which all of their actions are directed. When this happens, each agent conceives of the others' actions merely as constraints and opportunities. Bratman's conditions may initially appear be sufficient for shared intention because it may be natural for agents who meet his conditions to make use of the fact that there is a single outcome to which all of their actions are directed and to conceive of all their actions as parts of a single plan. Our counterex-

ample, the case of Beatrice and Baldrick, shows that although this may be natural, it is not necessary.

We identified a way to revise Bratman's account so as to avoid this sort of counterexample without departing substantially from his planning approach. The remedy we propose hinges on the possibility that an agent, in planning for an outcome, can sometimes plan both her own and other agents' actions. Further, in the right circumstances, two or more agents with the same task can each individually specify roles for all of the agents in planning for that task. Even where such planning occurs in parallel (and so need not involve no shared agency), it can sometimes rationally result in coordinated action. These reflections indicate that, in giving sufficient conditions for shared intention, we can appeal to the possibility that, in acting on a shared intention, each agent has a plan for all of their actions. This enables us to capture the idea that agents acting on a shared intention conceive of all of their actions as part of a single plan.

No amount of forming intentions about other's intentions and acting on knowledge of such intentions is sufficient, all by itself, for shared intention. In many or all cases, shared agency requires changing perspective to conceive of one's own and others' intentions and actions as parts of a single plan.

References

- Bratman, M. (1984). Two faces of intention. *The Philosophical Review*, 93(3), 375–405.
- Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review*, 101(2), 327–341.
- Bratman, M. (1993). Shared intention. *Ethics*, 104, 97–113.
- Bratman, M. (1997). I intend that we J. In R. Tuomela & G. Holmstrom-Hintikka (Eds.), *Contemporary Action Theory, Volume 2: Social Action*. Dordrecht: Kluwer. Reprinted in Bratman, M. (1999) *Faces of Intention*. Cambridge: Cambridge University Press (pp. 142–161).
- Bratman, M. (2009). Modest sociality and the distinctiveness of intention. *Philosophical Studies*, 144(1), 149–165.
- Call, J. (2009). Contrasting the social cognition of humans and nonhuman apes: The shared intentionality hypothesis. *Topics in Cognitive Science*, 1(2), 368–379.
- Carpenter, M. (2009). Just how joint is joint action in infancy? *Topics in Cognitive Science*, 1(2), 380–392.

- Gilbert, M. (2006). Rationality in collective action. *Philosophy of the Social Sciences*, 36(1), 3–17.
- Gilbert, M. P. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15, 1–14.
- 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.
- Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research*, 61(1), 1–31.
- Moran, R. (2001). *Authority and Estrangement: an essay on self-knowledge*. Princeton: Princeton University Press.
- Pears, D. (1971). Two problems about reasons for actions. In A. M. R. Binkley, R. Branaugh (Ed.), *Agent, Action and Reason* (pp. 128–153). Oxford: Oxford University Press.
- Rakoczy, H. (2006). Pretend play and the development of collective intentionality. *Cognitive Systems Research*, 7(2-3), 113–127.
- Roth, A. S. (2010). Shared agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2001 Edition)*.
- Searle, J. R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication* (pp. 90–105). Cambridge: Cambridge University Press. Reprinted in Searle, J. R. (2002) *Consciousness and Language*. Cambridge: Cambridge University Press (pp. 90–105).
- Tollefsen, D. (2005). Let's pretend: Children and joint action. *Philosophy of the Social Sciences*, 35(75), 74–97.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675–735.
- Tsai, J. C.-C., Sebanz, N., & Knoblich, G. (2011). The GROOP effect: Groups mimic group actions. *Cognition*, 118(1), 135–140.
- Velleman, D. (1985). Practical reflection. *The Philosophical Review*, 94(1), 33–61.
- Velleman, D. (2000). *The Possibility of Practical Reason*. Oxford: Oxford University Press.