CrossMark

Representing collective agency

Sergio Tenenbaum[1]

Abstract   This paper examines whether Bratman's succeeds in provides a reductive account of collective intention.

Michael Bratman's *Shared Agency*[1] provides a compelling account of collective intention and action; the account builds upon Bratman's earlier work on individual intention. The book has clear reductionist ambitions; Bratman argues that the resources of his theory of individual planning agency are sufficient to capture the emergence of what Bratman calls 'modest sociality'. Despite my enormous admiration for this book, I will be raising some concerns about Bratman's reductionist ambition; I'll argue it might be more difficult than *Shared Agency* suggests to capture a significant idea of collective agency using only the thrifty resources of Bratman's planning theory. A "non-reductionist" account might be what we need, and might not be as costly as *Shared Agency* suggests.[2]

The methodological advantages of reductionist views seem obvious; non-reductionist views charge an additional premium for explaining the same phenomena. But we should not exaggerate the price tag. There is obviously something strange about the idea of my bodily movements being controlled by an irreducible Group Mind. We gasp at the thought that certain manifestations of a

---

[1] (Bratman 2014). All parenthetical numbers refer to this book.

[2] Many of the concerns voiced here are anticipated in the book; in many ways, this piece is further reflection on concerns that Bratman himself raises in the book.

✉ Sergio Tenenbaum
    sergio.tenenbaum@gmail.com

[1] Department of Philosophy, University of Toronto, 170 St George St., 4th Floor, Toronto, ON M5R 2M8, Canada

person's agency are direct effects of some collective entity's states that bypass their individual attitudes or awareness. These kinds of cold-war-era-like nightmare scenarios should be metaphysically condemned, but non-reductionist theories of shared and collective agency rarely incur such extravagant ontological commitments. Instead, non-reductionist views are sceptical that any amount of individual intentions with "non-collective" contents could deliver an analysis, or account, of collective agency. Let us use the following locution for collective intentions in order to generate a schematic taxonomy of mild non-reductive views:

(CI) [We] [intend] [that X][3]

A non-reductive view might focus on the subject, attitude, or content expressed in (CI); none of these options needs to carry a particularly heavy metaphysical (or conceptual) burden. Let us start with the idea that the "we" is irreducible. Despite appearances, this need not amount to the postulation of an Orwellian collective mind. For instance, on Gilbert's view (or a variation of it), one might take the plural subject to be constituted by commitments that the agents in question jointly make.[4] The resources such an idea needs are not unlike the resources needed to account for various other joint commitments, such as contracts, business enterprises, wedding vows, etc. Another option is to take the attitudes in question to be irreducible; this would be something like Searle's "we-intentions". In such a case, the "we" in the locution does double duty. It denotes the multiple subjects of the attitudes (the subject "we" is understood distributively), and it modifies "intend" so as to make it denote a sub-species, or a different kind, of attitude. A we-intention is perhaps better represented as follows:

(CI*) I we-intend that X

Here we are postulating a new attitude, but an attitude that seems continuous with other attitudes we were already committed to (such as, of course, intentions).[5] Now I'm not very clear on what this attitude is, how many of these attitudes we'd need to make sense of collective intention (wouldn't we need a different one for each person I may collectively intend to do something with?), etc. However, I'm not sure we should be in principle very worried about its ontological or conceptual baggage. Compare it with a popular attitude whose postulation is still fresh off the shelves; namely, "alief".[6] Suppose one were to argue that the same phenomena that can be explained by postulating aliefs can be also explained by postulating a large set of highly complex beliefs that people hold in such a situation. Should this new view be preferred on grounds of parsimony? On the one hand, we have an explanation that relies on the postulation of an extra attitude, but an attitude that is not radically different from various other attitudes needed to explain other psychological phenomena. On the other hand, we can explain the phenomena instead with complex versions of familiar psychological attitudes that must be present in every

---

[3] Perhaps "We intend to ϕ" is the more basic locution. I'm not taking a stand on this issue.

[4] See (Gilbert 1990, 2009).

[5] See, for instance, (Searle 1990).

[6] (Szabo Gendler 2008).

instance of rather mundane phenomena. It is not clear that the distinction between quantitative and qualitative parsimony[7] gives us good reason to favour of the "reductionist" view of the alief phenomena.[8]

Finally, an irreducible content seems to be a particularly cheap option; we only introduce a new object of representation.[9] Now introducing an irreducible content invites a charge not only of irreducibility but also of circularity. Bratman considers whether such a charge can be levelled against his own view; he argues that at least as long as some basic cases can be explained in terms that do not presuppose a conception of collective agency, he can escape the charge of circularity.[10] But would a view that did not appeal to such basic cases fall prey to some kind of vicious circularity?

It is worth noting that this view doesn't imply that engaging in collective activity requires mastery of the concept "joint activity"; all one needs is access to the concept of an activity that is in fact irreducibly collective. Perhaps what I need in order to be capable of "modest sociality" is some concepts of activities that are essentially collective, such as conversing, dancing the tango, etc.

But one might think that as long as the theory employs the concept of a joint activity, whether explicitly or implicitly, in the content of the attitudes constitutive of collective intention or agency, it will be necessarily circular. But why? Such a theory would indeed imply that shared agency cannot emerge before the concept of shared agency (or of some forms of shared agency) emerge, but there is no contradiction or circularity in the assumption that this form of agency and its concept are necessarily coeval. Suppose a child is first going through the motions of ring-around-the-rosy without any awareness that she is playing a game or even that there are children around her. We can ask, paraphrasing Bratman paraphrasing Wittgenstein, what we need to add in order for the child to be engaged in social agency, to distinguish her agency from "a swarm of bees."[11] And it seems not too implausible that we need to add that the child is able to think of her activity as something that she does together with the other children.[12] Just as an illustration of how far our non-reductionist views are from spooky views, it might be worth considering a very simple view that has this form:

Simple view

(A) If X is ϕ-ing intentionally because$_{int}$[13] X and Y are (intend to be) ψ-ing jointly and Y is ϕ$_1$-ing intentionally because$_{int}$ X and Y are (intend to be) ψ-ing jointly then X and Y are (intending to be) ψ-ing jointly. [In the first

---

[7] (106).

[8] Similar things can be said, of course, about the other non-reductivist views.

[9] Kutz's notion participatory intention would be an example of a notion of an intention with an essentially collective content. See (Kutz 2000).

[10] More on this below.

[11] Bratman (10) quotes this example from (Petersson 2009).

[12] Perhaps this need not be through the acquisition of a concept of joint activity. The child might simply acquire the concept of 'ring-around-the-rosy' or 'this game.'

[13] The subscript specify that this is the "because" proper to intentional explanations or that this explanation is given in response to the Anscombean "why" question.

person case, the explanation X would explain some of her actions as following "I am ϕ-ing because we are (intend to be) ψ-ing together."]

Let us assume that we have a "good case"; namely a case in which X and Y share a common representation of "X" and "Y" (perhaps through essential indexicals) and of 'ψ-ing', etc. According to the simple view, shared intention and agency happen when the relevant individuals' actions must be explained in terms of their representation of a joint activity (either as the content of a prospective intention or as the content of an intention in action).

Of course, I am not trying to defend this view here; I'm only interested in the question of whether it is circular or ontologically rich. The simple view takes as basic intentions whose content would not be available for those agents incapable of representing some form of collective agency. But it does not introduce any content that is wildly foreign to our conceptual repertoire. Moreover, there is nothing obviously circular about such a view; the account explains shared agency and shared intention partly in terms of *representations* of shared intentions and activities but not partly in terms of itself. One might think that there is something strange that X could only exist if some representation of X existed. But arguably this is not unique to collective intentions. This seems to be true also of various forms of games, artefacts, commitments, money, legal institutions, etc.

So far I have just tried to caution against exaggerating the methodological advantages of reductionist views over views that rely on some non-reductive form of joint agency. But even if the alternative views turn out to be metaphysically modest, Bratman's theory of shared intention would still have a methodological advantage as long as he employs slimmer resources than his opponents in his account. So I would like to examine now if the theory can indeed lay claims to such advantages.

Bratman claims to be giving only sufficient conditions for shared intentions; thus Bratman allows that there might other forms of collective actions and intentions that are not covered by his analysis. But if this is the case, one might wonder why we should be so confident that the ultimate analysis of collective agency, the one that includes all the important forms of shared agency, will not introduce resources that are at least as committed to "discontinuities" from the individual cases as the alternative models of collective agency we encountered so far. I take it that Bratman's point is that once we have sufficient conditions for modest sociality, we have shown that sociality can indeed emerge from the planning theory proposed in his earlier writings. Thus, there'll be at least some version of modest sociality that is continuous with individual planning agency and that could not be better explained by alternative accounts; after all, any extra condition imposed by such accounts would be, *ex hypothesi*, in addition to Bratman's sufficient conditions. In fact anything, the existence of such "reductionist" sufficient conditions for modest sociality might give us hope that we'll find similarly reductive accounts for other cases.

However, I think that Bratman's claim for Ockhamite superiority on the basis of sufficient conditions alone is too quick. For one might think that the co-occurrence of these sufficient conditions is explained exactly by the joint commitment, representation of joint activity, or we-intention that the more expansive theory postulates. It is worth comparing a similar move that a behaviourist about pain

might make. Suppose that the behaviourist grants that the behavioural syndrome that she has used to define "pain" is indeed not necessary for pain; one can be in pain, for instance, while incapable or unwilling to move. However, she insists that she has given sufficient conditions for pain; whenever someone with no acting talents contorts herself in the right way, screams "I'm in pain", begs for help, etc., it is indeed true that she is in pain. Suppose she now claims that she has given an account of "modest pain" and that she can brandish her Ockham razor at least for these cases. It would seem fair for the non-behaviourist to reply that exactly those things we need to postulate in order to explain other cases of pain also explain why this syndrome occurs in these cases. Similarly, one might argue that it is in virtue of having a joint commitment to a complex action that agents engage in certain collective activities form interlocking intentions, etc.

Of course, this comparison is somewhat unfair. Our imagined behaviourist is probably accounting for a marginal case of pain while Bratman is arguably giving an account of some core, central cases of modest sociality. Although Bratman does not argue for it, Bratman's account of shared intentions might also provide necessary conditions for any, or virtually any, instance of collective agency. Moreover, even if we were to agree that in normal cases the sufficiency conditions are indeed explained by some deeper conditions of joint agency, should we not be able to find *some* counterexamples to the view in more atypical circumstances? In the remainder of the talk, I'll try, first, to raise some doubts about the extent of collective actions that clearly satisfy Bratman's conditions. I then proceed to present at least one case in which Bratman's conditions do obtain, but we do not have a genuine case of joint agency.

First, let us start by noting that for much joint activity, there seems to be no "pre-collective" version of the same activity. Some activities, such as dancing the tango, seem essentially collective.[14] Bratman suggests that as long as we can start from cases in which the activity can be described in "neutral" terms, we can allow that more complex forms of shared agency will emerge in which the shared activity is represented in essentially collective terms. But, in this context, I don't think this is a minor concession. These kinds of shared activity are not particularly complex or rare; they range from playing ring-around-the-rosy to conversing or idle chatting. Once, while in the parrot enclosure of a bird sanctuary I uttered the usual "hello" in the hope that a gray parrot would answer back. Amazingly the parrot did not only repeat what I said, but also added some other exclamation. I repeated what the bird had added and tried to keep it going with a new bit of content. The bird did the same. I went on like this for a few rounds until I realized that the bird had been completely silent; the words have come from another man on the other side of the exhibit who was fooled by my utterances just as I was fooled by his. Despite the surface similarity, my unwitting partner and I were not engaged in a pre-collective version of conversation; intending to have a real conversation is not intending to have this kind of interaction in a more mutually responsive manner.

However these simple, "essentially" collective interactions cannot be represented as sharing the intentions that are constitutive of shared agency in Bratman's model, at

---

14 This is Bratman's own example.

least not without adding to the representation something that is essentially collective. It would be obviously no help to allow the agent to represent explicitly Bratman's sufficient conditions as long as we do not introduce as part of the content of the intentions the irreducibly collective activity. Of course, Bratman could allow that the content of the intention to be irreducibly collective in these cases. However, my suspicion is that once we allow irreducibly collective contents, most of the other conditions turn out to be moot. If I am intending to be engaged in a joint activity, we do not need to postulate that I also intend that my subplans mesh with my partner's subplans; whatever commitments, dispositions, or norms follow from the latter intention, would already be consequences of the irreducibly collective intention.[15] Moreover, this approach seems to have the advantage of getting the relevant explanatory relations right. Let us take the case of going for a walk together. It seems that what *explains* our having intentions that our subplans mesh is our shared interest in taking a walk together; ensuring that our subplans mesh is the *means* by which we ensure that we successfully engage in this joint activity. But this explanation is not available to Bratman;[16] Bratman needs to say that we are walking together (partly) *in virtue of* our intending our subplans to mesh, and this intention would probably be explained in terms of meshing subplans being means to something else or by the fact that we surprisingly desire our subplans to mesh for its own sake.

Moreover the conditions of interlocking intentions seem in the context of such simple cases of shared agency to be overly demanding. There is, first, the question of whether children or unsophisticated agents have all those intentions every time they engage in conversation or in a game of ring-round-the-rosy. Bratman addresses this concern by pointing out that, the intentions in questions can be "only tacit or implicit", or even in some cases "exercise [merely] virtual control" (104). It is always difficult to see how understand these implicit or "standby" mental states. Let us take, for instance, the relevant instance of (iii) (84):

(iii$_{rrr}$)Each child intends: we each play ring-around-the-rosy by way of sub-plans of each of our intentions in favour of playing ring-around-rosy that mesh with each other.

This seems a rather complex thought for the average ring-around-the-rosy aficionado. But do the children have such intentions implicitly, tacitly, or by exercising "virtual control"? Doubtless a child needs to have the capacity to adjust her behaviour appropriately if she's going to succeed in the game, but this is not necessarily different from the capacity that she would exhibit in adjusting to the movement of inanimate objects; an intention with the particular content of (iii$_{rrr}$) is explanatorily idle. Moreover, an intention comes with dispositions, rational requirements,[17] etc. that might be simply absent in such cases. Our children might be very particular, unwilling to make adjustments to their plans, etc. without thereby

---

[15] It is worth noting that this is not a problem for Gilbert, for instance. The content in question would simply be the representation of something that essentially involves joint commitments.

[16] At least not in the "basic cases".

[17] At least on Bratman's own view.

exhibiting any incoherence in their intentions to play the game. And it is not just child's play that seems to proceed without such complex intentions; similar things could be said about casual chat with a stranger on the bus, as well as other "small" joint activities.

Finally, the condition of common knowledge would also be too strong as a necessary condition. Suppose I am a human tester in a Turing machine contest. So I start "conversations" with various partners, whom in the end I'll need to classify as either human or machine. As conversation progresses with one of them, let's call him "Hal", we really hit it off. We become such quick friends that we decide we must hang out and spend some quality time together; we start to plan a trip together. I make hotel reservations; Hal checks flights for us. I rent a car while Hal makes us reservations at romantic restaurants in the area. We go back and forth about dates, locations, etc. ensuring that the various plans we're making mesh. Everything goes swimmingly, but unfortunately I know how good the programs are. For all I know, Hal is not making any reservations, intending to take a trip, feeling any vibes. Hal is quite possibly just a computer program executing sub-routines that will give answers that will sound reasonable to a human mind. So I don't know that Hal intends anything, and yet if Hal turns out to be human, it'll be true that we just arranged a joint trip together (let alone that we conversed, coordinated our reservations, etc.), and that we shared intentions to do so.[18]

But even if the claim of sufficiency were enough to give Bratman the methodological pole position, I have doubts that the conditions are indeed sufficient.

Let us suppose Larry has an inordinate amount of money coming to him. But the bank will give Larry the money only if he shows up at the bank with Mary so that a bank teller can confirm that Mary is alive. Unfortunately, as Larry knows, Mary would prefer him to be penniless, so she's not willing to go the bank with Larry simply to help him. However, as luck would have it, Larry finds out that Mary is in the perfectly symmetrical predicament. As one would expect, Larry feels the same spitefulness toward Mary. Their mutual hatred runs deep but not as deep as their self-love; each would rather have the money rather than let the other suffer in poverty. Unfortunately, more powerful than their mutual hatred or their inflated self-love is their sense of dignity. Neither would stoop to ask the other a favour or propose a truce or an agreement. They all know all this and are thus incapable of retrieving their respective fortunes. One day Larry is walking and he sees Mary; he immediately realizes that if he walks towards the bank, she'll follow him there in the hopes she'll get the money. Mary is hit by the same thought. They immediately notice that both would have had the same thought, and that they both realize that the other would have had noticed that both had the same thought, etc. They walk towards the bank, each carefully monitoring that the other is going. This seems to be a case in which no sociality has emerged—of either the modest or the expansive kind. In fact, this is a rather pure case of strategic interaction. Yet all of Bratman's conditions hold.

---

[18] I can't dwell on this here, but I think the competing views can account for such a case.

Of course, we can imagine a version of this case that there would be so much cooperation, or at least disposition to cooperate, that this would be arguably a case of shared agency. If they were willing to help each other every time they stumbled, or to communicate on the more effective way to get to the bank, etc. we would have a case of shared agency. But they could have the minimal dispositions required for counting as having the intention "Larry (Mary) and I go to the bank"; they could, for instance, each be disposed to avoid paths that would obscure them from the other's view or that would lead the other to collide with an obstacle on their way to the bank, but not much else. At any rate, this is a case in which Bratman's verdict parts company with the verdict of the competing theories. There is no "we-intention" here, no representation of an essentially joint activity, and certainly no joint commitment. I find myself here siding with the opposing views; my fear is that such cases might show that reductionist accounts cannot fully capture the "sharedness" of shared agency.

In sum, I think that Bratman's claims to relative parsimony are overstated. The opponents are introducing entities and concepts that are not essentially different from the ones we already have available anyway; the contrast between quantitative and qualitative parsimony is not so clear-cut in this debate. Moreover if we are serious about the account providing only sufficient conditions, the extra resources used by non-reductionists might be needed elsewhere anyway, and in turn be able to explain why Bratman's account provides sufficient conditions in a subset of cases of joint agency. I also argued that accounting for certain relatively simple cases of joint action might already require more robust conceptions of joint intention and agency. Finally, I tried to call into question whether Bratman's radical reductionism can provide sufficient conditions for modest sociality.

These comments are, of course, not intended as refutations. Bratman has provided us with an original and compelling understanding of shared agency; one can hardly scratch the surface in such a brief paper, let alone do much more than raise a few relatively small challenges to the view. In other words, my aim here was very modest disagreement.

References

Bratman, M. (2014). *Shared agency*. Oxford: Oxford University Press.

Gendler, T. S. (2008). Alief and belief. *The Journal of Philosophy, 105*(10), 634–663.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies In Philosophy, 15*(1), 1–14.

Gilbert, M. (2009). Shared intention and personal intention. *Philosophical Studies, 144*, 167–187.

Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research, 61*(1), 1–31.

Petersson, B. (2009). Collectivity and Circularity. In H.-B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of sharedness—Essays on collective intentionality* (pp. 145–164). Berlin: Ontos Verlag.

Searle, J. (1990). Collective Intentions and Actions. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communications* (pp. 401–415). Cambridge: MIT Press.