

Review of Michael E. Bratman's *Shared Agency: A Planning Theory of Acting Together* (Oxford University Press 2014)¹

If you have ever had to move house, you will know this: the worst part is the sofa. You cannot do it alone. Nor will it be enough for me to just lift one end waiting for you to lift the other. We will have to work *together* to get the job done. If spaces are tight, we will even have to find a practical solution to a tantalizing mathematical puzzle: the moving sofa problem.

Joint actions like that are part and parcel of everyday life. But what exactly is special about acting *together*? After all, the actions of two strangers also depend on one another when one exits and the other enters through a revolving door, when they happen to walk side-by-side along a forest path, or when they exchange blows in a pub brawl.

The problem is that two patterns of social behavior might look identical, even though one is a case of joint action and the other is not. There need not be an observable difference between the movements of old friends taking their morning constitutional together, on the one hand, and those of two strangers merely trodding a parallel path, on the other. This suggests that “jointness,” unlike meaning, is in the head. So what sort of mental attitude grounds “jointness”? The above examples point to the central role of intentions. Friends’ movements are guided by their shared intention to take a stroll together, whereas in the case of strangers the intention of each is limited to getting from *A* to *B* along the path.

Summarizing and elaborating his already highly influential *oeuvre*, in this book Michael Bratman gives us his definitive account of shared agency. Bratman accepts that joint actions must be categorically distinguished from other types of social interactions which may be coordinated or co-dependent but do not involve shared agency. This is an idea he shares with other members of the “Big Five” of the collective intentionality and social ontology literature: Margaret Gilbert, Philip Pettit, John Searle, and Raimo Tuomela.

What is original about Bratman’s approach is his account of intentions underlying joint actions. The key is the *continuity thesis*. This states that the right understanding of *individual* intentional agency is capable of explaining shared agency. The continuity thesis is crucial for two reasons. First, it allows Bratman to build his theory of shared agency on top of his theory of individual intentionality – the so-called planning theory of agency. Second, the continuity thesis implies that no novel conceptual, metaphysical or normative building blocks have to be introduced in order to explain what distinguishes joint actions from strategic or coordinated social interactions. This is important because Bratman recommends the theory over its rivals precisely on the strength of its

¹ Forthcoming in *Ethical Theory and Moral Practice*, online first: March 2015. DOI 10.1007/s10677-015-9585-5.

parsimony. He claims that his theory of shared agency can reduce, for example, Searle's "we-intention" or Gilbert's "joint commitment" by "providing reductive sufficient conditions for the joint-ness and shared-ness at work" (155).

Spelling out the details of this reductive account takes up the bulk of the book. Bratman aims to show that individual agents will share their intention to *J* if (i) their individual intentions in favour of *J* are appropriately interrelated, (ii) have certain beliefs about the efficacy and persistence of the intentions of each in favour of *J*, (iii) and their having these intentions and beliefs is a matter of common knowledge. If, furthermore, when seeking to satisfy this shared intention the individual agents exhibit (iv) mutual responsiveness in how they actually form the various sub-intentions, and how they perform their own contributions required to *J*, then *J* is rightly called a shared intentional activity. In short, two (or more) individual planning agents act jointly if their relevant intentions and contributory actions are appropriately responsive to the intentions and contributory actions of the other(s), and they are aware of this connection.

Bratman responds in considerable detail to several objections. Three of these objections have already spawned intricate side-debates in the literature. The first is the circularity objection. Bratman argues that all participating agents should have the intention "I intend that we *J*" (and not just "I intend to contribute to our *J*-ing"). But if "...we *J*," is to be read as "*J*-ing together" in sense defined above, then both the *analysandum* and the *analysans* refer to shared agency. So we have not succeeded in reducing or even explaining shared agency.

The second objection is that intention is not the kind of thing that can be shared. Many believe, for example, that one can only intend one's own actions: *I* cannot make up *your* mind. The third objection is that Bratman's account is too demanding. After all, children and animals cooperate successfully, even though they lack the cognitive capacities required by Bratman for shared agency.

In addition to answering these standard objections, Bratman spends considerable time responding to rival theories. Since Bratman offers sufficient rather than necessary conditions of shared agency (36), he has a bone to pick mainly with those who think that his conditions are insufficient. Bratman seeks to demonstrate that his reductive approach can avoid the drawbacks of theoretical alternatives. First, he thinks it manages to capture the inherently interpersonal character of shared agency, thereby steering clear of what many see as an implausible implication (even *reductio*) of Searle's conception, namely that even a single individual, and perhaps even a brain-in-a-vat can be ascribed a "we-intention". Second, Bratman believes to have improved on Gilbert's approach which explains joint action in terms of mutual obligations and joint commitments. He shows *how* joint action is possible without mutual obligations and joint commitments.

Finally, Bratman makes an important reply to Tuomela's account of sociality. Participants, Bratman insists, may have quite different reasons for contributing to shared agency. The "pervasiveness of partiality in our sociality" helps to

explain how shared agency is possible even under circumstances of disagreement and dissent – an issue that looms large in pluralistic and multicultural liberal societies.

Rather than trying to assess Bratman's replies, I would like to take up two topics (and mention a third) which I found wanting in the book, and close with an objection. Let me begin with the missing topics.

With the exception of a few short passages, Bratman avoids the problem of moral (and non-moral) responsibility. This is a pity. The issue of collective agency has always been approached from two different directions. The action-theoretical angle, dominating this book, focuses on the issue whether certain ideas about individual agency can be extended to groups. However, the normative angle is just as important. Many have wondered whether groups *qua* groups can be responsible in a non-distributive sense, i.e., whether the ascription of responsibility to a group can mean more than a summation of the responsibility of individual members. But even when responsibility-ascription is understood distributively, it has been suggested (Petersson 2008) that we sometimes have to proceed *top-down* when allocating responsibility, i.e., hold an individual responsible not on the basis of her contributory action, but rather hold her responsible for the group's action as a whole given her participation in shared agency. The promise of such an approach is that it could deal with cases in which individual contributions do not seem to be culpable when considered separately from a *bottom-up* perspective – not culpable because the contribution is marginal or because it is individually rational (as in tragedy-of-the-commons type cases). Exploring the implications of the reductive account for the distribution of responsibility could also be useful in testing the applicability of the theory to more controversial cases, especially those in which shared agency is constituted by strongly unequal partnerships involving coercion or deception.

Causation is another topic that is mostly avoided. This is also a little surprising in light of important recent attempts to ascribe irreducible causal agency to groups. What makes these attempts relevant is that if it can be shown that the causal agency of some groups supervenes on but does not reduce to the agency of participating individuals, then it becomes tempting to see these groups as agents in a stronger sense than what the reductive account appears to allow. Of course, Bratman can reply once again that his aim was only to provide sufficient conditions for shared intentional agency. Moreover, according to most collectivists only larger groups with lasting institutional structures in place qualify as robust collective agents (e.g., corporations). By contrast, Bratman's theory of shared agency is emphatically of *modest* sociality. It applies to small-scale and often transient partnerships. But the problem is not just the failure to engage with what many see as an exciting part of the story about collective action. A more pressing issue is that even some cases of shared agency could actually involve group agency in a more robustly collectivist sense. If so, these instances of shared agency may not be reducible to the planning agency of individual contributors.

I can only mention in passing another problematic issue that a full critical appraisal would certainly have to discuss: the condition of mutual knowledge required for shared agency. Bratman does not analyze this condition, although it poses a serious challenge to reductionism about shared agency (Schmid 2005).

Finally, an objection. A point frequently repeated in the book is that as long as individual agents seek to cooperate and each of them individually accepts individualistic norms of rationality (28) – in particular those of consistency, agglomeration, coherence and stability (27) – these individual agents, when acting jointly, will also conform collectively to what Bratman calls “social rationality norms,” i.e., “norms of social agglomeration, social consistency, social coherence, and social stability” (87). It follows, according to Bratman, that violations of social rationality norms “normally consist of a violation of associated norms of individual planning agency” (87).

This claim sits rather uneasily with central findings of game theory, on the one hand, and social choice theory, on the other. The former tells us that in many quite ordinary types of situation individual and collective rationality may be *necessarily* at loggerheads with each other. The latter shows that individually rational forms of behavior or attitudes can lose some of their rationally desirable properties when aggregated. This is the upshot of Condorcet’s paradox, which was generalized by Arrow in his famous “impossibility” theorem. In light of these findings, it is unclear whether Bratman is entitled to positing a straightforward link between individual rationality and collective rationality.

The relevance of this problem is not limited to special cases. It casts doubt on Bratman’s claim that as long as contributing agents pursue individualistic norms of rationality “it should be possible to agglomerate relevant intentions into a larger social plan that is consistent, that in a timely way adequately specifies relevant means and preliminary steps, and that is associated with appropriately stable social psychological structures” (27-8). *Pace* Bratman, what we learn from game theory and social choice theory is that there is a conceptual and normative discontinuity between individual planning agency and the “distinctive, rational normativity of modest sociality” (105). This does not mean that shared agency is impossible, but it does mean that it can be quite difficult to realize.

There appear to be two different solutions to this problem. The first is more collectivistic than Bratman’s approach, the second is (even) less collectivistic than his. The *more* collectivistic option is summed up by Tollefsen: “[...] groups rather than simply the individuals that comprise them are subject to the norms of rationality and form a distinct locus of power and responsibility. This makes our application of the intentional idiom to groups intelligible and suggests that groups are literally intentional agents.” (Tollefsen 2002a, 34; see also List & Pettit 2011; Rovane 2004). Taking this option entails that the connection between “group agent-hood and group-subject-hood” (see 189n22) might be tighter than Bratman thinks. A further implication is that the partiality of sociality may not be as pervasive as Bratman suggests. In some cases, shared agency may be impossible unless individual agents are willing to be guided by group reasons.

The second option is more individualistic than Bratman's own. Unlike Bratman and like collectivists, this approach recognizes the possibility of a discrepancy between individual and collective rationality. Unlike collectivists, however, it holds that collective practical irrationality should not be described as the failure of a collectively intentional agent (for some arguments in favour of this option, see Szigeti 2014). On this approach, either an individual person is to be blamed for practical collective irrationality or nobody is.

References

List, C. & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press

Petersson, B. (2008). Collective omissions and responsibility. *Philosophical Papers* 37: 243-261

Rovane, C. (2004). Rationality and Persons. In A. Mele & P. Rawling (eds.), *Oxford Handbook of Rationality*. New York: Oxford University Press, 320-342

Schmid, H.B. (2005). *Wir-Intentionalität*. Freiburg i.B.: Karl Alber

Szigeti, A. (2014). Why Change the Subject? On Collective Epistemic Agency. *Review of Philosophy and Psychology (online first)*: 1-22

Tollefsen, D. (2002). Collective intentionality and the social sciences. *Philosophy of the Social Sciences* 32: 25-50