

Michael E. Bratman, *Shared and Institutional Agency – Toward a Planning Theory of Human Practical Organization*, Oxford University Press 2022, 288pp., \$29.95 (pbk), ISBN 9780197580905

Reviewed by Abraham Roth (Ohio State University)

In his latest book, Michael Bratman works toward an account of institutions and institutional agency, starting from his well-known planning theory of intention and the important work he's done more recently on shared agency. He also draws on H.L.A. Hart's influential views about the law, in particular his account of social rules. The book ends with a discussion of the intentionality of groups. There he articulates a form of intention decoupled from reasons. He argues that institutions can in this sense possess intentions and be genuine intentional agents, even though they are not rational subjects in the way that individual agents are. The book is a clear, intricate, and insightful work that will be a crucial reference point for theorizing about practical reasoning, shared agency, collective intentionality, and the metaphysics of the social world.

Bratman sees this project as one of scaling up (21) his account of small-scale joint action into a theory about the action of large groups or institutions (respectively, the 'Shared Agency' and the 'Institutional Agency' of the book title). Some (e.g. Searle) take the theoretical resources developed to account for what it is for several individuals to act together and apply them directly at a large scale (e.g. to give a story about money or borders, or government bodies and corporations). Bratman does place his account of shared agency at the core of the story about institutional agency ("central claim of this book" 9), but he appreciates that scaling up won't be straightforward.

Accounts of (small scale) shared agency seek to characterize the distinctive ways in which individuals are related to one another when they act together, as opposed to when they are acting on their own or interacting with one another strategically. Bratman's account (defended in a number of articles and in his 2014) sees this relationship as involving each participant bringing what the others do within the scope of their will. More specifically, each participant has an individual planning intention of the form *I intend that we J*, thereby intending what every participant does. Some find intentions of this form problematic. For example, intending something settles the practical question of what to do. But it's not obvious that one can settle what *we* do in anything like the manner in which one can settle what one does on one's own. And even if it is possible, one might wonder how this can be a part of the story of acting together rather than a picture of one individual manipulating and undermining the agency of another. But Bratman has argued that under the right conditions, each individual will be able to form their own relevant planning intention, and this interpersonal nexus of intentions (what Bratman calls a *shared intention*) can account for much of what is distinctive about shared agency – in particular, how participants in joint action relate to one another. For instance, joint action exhibits what we might call practical intersubjectivity: participants intend, plan, and act in ways that cohere with those of fellow participants. Bratman contends that his notion of shared intention enables us to explain the *interpersonal* consistency and coherence of intentions

characteristic of shared agency in terms of the *intrapersonal* norms associated with individual planning intentions (13,17).¹

But the conditions that suffice for shared intention amongst a couple or a few agents generally won't hold at larger scales (22). So Bratman carefully develops a set of conditions sufficient for a variety of shared intention more appropriate for larger scale joint action (48). This is in keeping with his idea that the nexus of individual intentions constituting shared intention of one form or another is key to understanding not only shared agency but important aspects of institutional agency as well. At the same time, however, Bratman maintains that it would be a mistake to think that institutional agency is a matter of shared agency writ large (xviii, 33). Not everyone in an institution is acting together. Rather, institutions usually exhibit a modular structure. Individuals constituting a subgroup within an institution might act together and share an intention, but individuals across the whole of the institution often do not. (Contrast Ludwig's important discussion (2017) where he invokes the notion of proxy agency to maintain that all members of an institution are agents of whatever the institution does.)

But if not everyone within an institution is party to a shared intention, what unifies the various elements of an institution – the modules within which individuals do share intentions – into a single entity at a time and over time? And how are we to characterize the relationship between members of an institution who don't necessarily share an intention? Bratman thinks organized institutions are typically rule governed and rule guided (33). So his strategy for addressing these questions invokes an understanding of social rules (119). The rough picture is one that characterizes an institution in terms of a web of social rules specifying the function of its various parts, where the functioning and output of those parts is governed by overarching rules of procedure that resolve conflicts that might arise between the operations and outputs of the different parts (115).

Bratman thinks that Hart's characterization of social rules is particularly helpful in capturing the modular structure we find in organized institutions (33-34). He sees a Hart style theory of social rules as "adding range to the planning theory" (xviii). Now, social rules are not mere abstract specifications of appropriate conduct. Like regularities of conduct, social rules are themselves realized in an "interpersonal causal order" (34). But, for Hart, they are more than regularities in that they have an internal aspect. Individuals *take* some patterns of behavior in appropriate circumstances as a standard to be followed and this is part of the explanation of the regularity. These standards are the basis for criticism of oneself and others for non-compliance, and underwrite interpersonal guidance and demands for conformity (39, 43). Bratman understands this psychological attitude (the aforementioned taking) in terms of his planning theory of intention and practical rationality: we have an intention or policy to conform to the relevant regularity. The rationality of this interpersonal criticism and guidance is not so much a matter of there being a substantive reason to comply with the standard, since the social rule and the institution in question might for example be morally bankrupt (52). The interpersonal criticism and guidance are instead grounded in the rationality governing the coherence and consistency of attitudes, especially those of intention and plans. Bratman sees Hart's characterization of social rules as a kind of "design specification" to be realized by something like his picture of shared

¹ See Gilbert (2009 and many other works) and Roth (2004, 2014) for alternative proposals concerning the relation between participants in joint action.

intention or policies (suitably modified for larger contexts). So the interpersonal criticism or guidance is, at the fundamental level, a demand that one act in a way that rationally coheres with the shared intention – specifically, with the element of the shared intention that is one’s own intention to take some pattern as a standard.²

At this point, one might wonder how social rules could as Bratman says “add range to the planning theory” if the rules themselves are realized and hence constrained by shared intentions, i.e. by the nexus of individual planning intentions. And if institutions are individuated at least in part by some overall rules of procedure realized by shared intention/policy, isn’t there a sense in which institutional agency is a form of shared agency writ large? If a rule of procedure ultimately has authority over the entirety of the institution, mustn’t everyone in the institution participate in the shared policy/intention that is the realization of the procedural rule? But didn’t Bratman insist that institutional agency was not shared agency writ large?

It might be that Bratman is making an exception for rules of procedure for an institution, in which case the imposition of a rational order over the operations of an institution *would* be an exercise of shared agency writ large. But an alternative understanding more in keeping with the modular picture Bratman espouses draws on what he describes as the kernel/penumbra structure of social rules (66). A social rule is implemented by agents who have the relevant shared intention or policy. These individuals constitute the kernel for the social rule. But the rule can extend beyond them to apply to more alienated individuals on the periphery who are not party to the shared intention or policy. Such individuals might be under rational pressure to conform; they might have strategic reasons or reasons of convenience to conform to rules widely shared; the rule in question might serve as a focal point for coordination amongst a host of individuals not party to the underlying shared intention. The idea, then, is that the rule might be deployed in interpersonal rational engagement well beyond those who are party to the underlying shared intention or policy.

This idea should also apply to the rules of procedure governing the institution. Some individuals might be party to the shared intention or policy implementing whatever rules of procedure might be in force. But the rule of procedure might extend beyond this kernel to a periphery that includes other individuals and groups within an institution governed by those rules. The procedural rules are implemented by a governing body which is just another module within the institution, with everyone else falling within the penumbra of those rules. This picture appears to reconcile the idea that rules of procedure might govern and relate the relevant modules of an institution without necessarily invoking an exceptional shared intention writ large across the entirety of the institution.

The kernel/penumbra structure represents a substantive departure from Bratman’s original picture of the shared intention as underwriting the special way in which participants are related to one another. Participants in shared agency are subject to a distinctive kind of rational critique and guidance underwritten by the nexus of individual planning intentions that is the shared intention. But we’ve seen that Bratman thinks this form of interpersonal rational guidance and

² Intention is a committal state, and has a sort of inertia to it so that one is settled on a course of action. Moreover, if one’s conduct is coordinated with others who similarly intend, there might be even more reason for the intention to be stable and embody a commitment. But it is a good question whether a social psychological structure of such individual intentions will suffice to capture the force and authority that at least some rules seem to have over us.

critique has only a limited reach within institutions and does not bear on the individuals who are on the periphery of some social rule or institution. Whatever guidance and criticism those on the periphery are subject to is more strategic in nature – a matter of rationally coping with the institution within which such individuals find themselves embedded (68ff). It is interesting that the similar forms of interpersonal criticism and guidance to which individuals in the kernel and penumbra are subject can be grounded in such different ways – a difference which prompted theorizing about distinctive nature of shared agency in the first place. This might reflect Bratman’s appreciation of the complexities of institutional phenomena and an admirable restraint on his part in not wanting to overextend the resources of the planning theory of shared intention. But if Bratman concedes that much of the interrelatedness of individuals within an institution shouldn’t be explained in terms of the distinctive way in which individuals are related in shared agency, then one might be tempted to develop a more unified account that accords little significance to shared intention. Of course, shared agency might play an historical role in the establishment of institutions, but Bratman’s interest is not in the etiology of institutions, but in providing a construction of the “conceptual, psychological, metaphysical, and normative structure” to illuminate the social world. If the project is that of rational reconstruction, then why not eliminate the reference to shared intention for the sake of a more unified theory?

A related issue arises for Bratman concerning the persistence of institutions. An institution at one time might be built around a certain population of individuals who share the relevant intentions or policies underwriting some social rules. Similar sets of rules might be held by other perhaps overlapping populations at subsequent times, and there might be a causal explanation for this. Moreover, the rules in the earlier institution might target individuals in subsequent populations. In effect, those in the subsequent populations would be in the periphery of the earlier institutional stage even if they are within the kernel of a subsequent stage. In this way, Bratman develops a Lockean picture of the persistence of institutions over time (75ff). But none of this would seem to ensure that the kind of interrelatedness of individuals that underwrites interpersonal criticism and guidance in shared agency will extend across time with a change of populations. Whatever interpersonal (in this case intergenerational) rational guidance and criticism there may be within an institution that persists over time with a change of population, it would be of a very different sort from the planning rationality that one finds in shared agency, which for Bratman doesn’t allow for a change in individuals. Of course, what might be a worry for the view could instead be a realistic take on Bratman’s part regarding the possibility of intergenerational conflict. The thought would be that we shouldn’t expect that individuals from different generations within an institution should be related in anything like the way that individuals in joint action are, and we shouldn’t expect that the social rules of a previous generation have any authority over the current generation. It might only appear that way given the similarity of these rules to those underwritten by the shared intention of the current generation. The reality suggested by the view, however, might be better depicted as a matter of the current generation coping with rather than participating in the institutions of the previous generation.

Organized institutions typically are not integrated in such a way as to have a shared intention across all individuals within it. In perhaps the most provocative part of the book, Bratman argues that we can nevertheless make sense of an institution *itself* having intentions and acting intentionally (xx). Bratman specifies what sort of actions taken by individuals or groups within an institution count as outputs of the institution itself (chapters 5 and 6). He then argues in chapter 7 (137) that this institutional output can have a downstream functional role similar to that of

intention within an individual's psychology, such as securing means to promote the outcome, preventing pursuit of conflicting goals, etc. Bratman thinks that this similarity in functional role suffices for the output genuinely to count as an intention, despite lacking the rich backdrop of rationally constrained mental states that we always find in the individual case (Chapter 9). In this respect Bratman's approach differs substantially from those of other philosophers such as Tollefsen, Rovane, List and Pettit who also take seriously the ascription of mental states to institutions or groups.

It could very well be that having a particular functional role will suffice to distinguish intention from other mental states within an individual's psychology. However, it's unclear whether some social phenomenon that plays a similar role within a community should also count as the mental state of intention. Functional roles might distinguish one mental state type from another, but there might be more to mental states than functional roles. In any case, it is arguably only against a rich cognitive background of other mental states of the subject that a state with a certain functional role can count as an intention. And the backdrop of other mental states of the institution appears to be sparse if it exists at all. For Bratman, however, this merely shows that institutional intentions are a distinct species of intention. I think that Bratman is untroubled by this worry about the status of institutional intention qua intention because there is in fact a substantive cognitive background for institutional intentions. Although the institution itself lacks a substantive holistic web of mental states, a cognitive backdrop is nevertheless present in the form of a rich distributed web of attitudes of participating individuals (153-154). This response is fine as far as it goes, but one thing that the cognitive background secures in the individual case is a handle on the reason for action that makes sense of intending that action. Whereas, in the institutional case, the attitudes of participating individuals need not in any straightforward sense yield a coherent or unified reason for the intention and for performing the corresponding action (157). People might collaborate for all sorts of different and incompatible reasons. Given the possibility of disagreement, it could be in the interest of the institution simply not to take a stand as to why or for what reason it intends the action in question. Correspondingly, Bratman regards it as a feature rather than a bug that his characterization of institutional intention doesn't require a connection to something of central importance to philosophers of action such as Davidson and Anscombe – namely, the intelligibility from the agential point of view that reasons confer on action. Bratman's challenge to the Davidson/Anscombe orthodoxy is interestingly motivated by certain aspects of how institutions work. This important book should spark debate on this as well as a host of other issues.³

References

- Bratman, Michael E. (2014). *Shared Agency: A Planning Theory of Acting Together*. New York: Oxford University Press.
- Gilbert, Margaret (2009). "Shared Intention and Personal Intentions." *Philosophical Studies* 144: 167–87.
- Hart, H. L. A. (2012). *The Concept of Law*, 3rd ed. Oxford: Oxford University Press.
- List, Christian, and Philip Pettit (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.

³ Thanks to Olle Blomberg for very helpful conversation and comments.

- Ludwig, Kirk (2017). *From Plural to Institutional Agency: Collective Action volume II*. Oxford: Oxford University Press.
- Pettit, Philip (2003). "Groups with Minds of Their Own." In Frederick Schmitt, ed, *Social Metaphysics*, 167–93. Lanham, MD: Rowman & Littlefield.
- Roth, Abraham Sesshu (2004). "Shared Agency and Contralateral Commitments." *The Philosophical Review* 113: 359–410.
- Roth, Abraham Sesshu (2014). "Prediction, Authority, and Entitlement in Shared Activity." *Noûs* 48: 626–52.
- Rovane, Carol (1998). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.
- Searle, John (1995). *The Construction of Social Reality*. New York: The Free Press.
- Tollefsen, Deborah Perron (2015). *Groups as Agents*. Cambridge: Polity Press.