

IS COLLECTIVE INTENTIONALITY REALLY PRIMITIVE?

© Élisabeth Pacherie

PLEASE DO NOT QUOTE WITHOUT PERMISSION

1. Introduction

The purpose of Searle's book, *The Construction of Social Reality*, is to provide a conceptual foundation for the social sciences. The project is mainly ontological. Searle is a convinced monist. He believes that there is but one world, whose most fundamental features are described by the natural sciences. *Prima facie*, it is not obvious how certain phenomena of central concern to human beings in general and to philosophers in particular can be accommodated within a unitary physicalist ontology. Much of Searle's earlier philosophical work has been devoted to investigating how this could be done for linguistic and mental phenomena. *The Construction of Social Reality* extends the investigation to social phenomena and attempts to develop a general theory of the ontology of social facts and institutions. Searle sets out to show how complex social phenomena can be accounted for using a limited number of conceptual tools. His theoretical apparatus comprises four main elements. The first three — the assignment of agentive functions, collective intentionality, and constitutive rules — are introduced and described in the first chapter of the book. The fourth element, what Searle calls the Background, is introduced in chapter VI to explain the causal functioning of institutional structures.

In this paper, I shall be concerned mainly with collective intentionality and the account Searle gives of this notion. I shall also discuss the Background insofar as collective intentionality is taken to presuppose certain specific Background capacities. Searle argues that collective intentionality is irreducible to individual intentional behavior and should be considered as a biologically primitive phenomenon. Although I agree that collective intentionality is irreducible to individual intentionality in a certain sense, I do not think that collective intentionality must be either reducible to individual intentionality or biologically primitive. I will start by discussing Searle's account of collective intentionality and how it is constrained by his further commitments to individualism and to internalism. I will also indicate what I find the shortcomings of this account to be (section 2 and 3). The fourth section will outline an alternative account of collective intentionality which, I hope, overcomes some of these shortcomings without being subject to the objections Searle addresses to reductive analyses. In the fifth section, I will discuss the capacities needed to sustain collective intentionality according to this alternative account. In the last section, I will briefly examine what becomes of Searle's requirements of individualism and internalism on this account.

2. Searlian intuitions

The notion of collective intentionality plays a crucial role in Searle's account of social reality. Of the three conceptual tools introduced in the first chapter of the book, it may be deemed the most central. It is indeed a defining feature of social facts since, for Searle (p. 26), any fact involving collective intentionality qualifies as a social fact. By contrast, agentive function is not criterial of social facts since there can be singular as well as collective assignment of agentive functions on objects. As for constitutive rules, they are involved only in one special subclass of social facts, namely institutional facts. Despite its centrality, collective intentionality is characterized rather cursorily in *The Construction of Social Reality* and, for a more detailed account of this notion, one must revert to an earlier paper of Searle's (Searle, 1990).

Searle's view of collective intentionality is based on a twofold intuition. The first part of the intuition — that there is collective intentional behavior as distinct from individual intentional behavior — is hardly controversial. Searle offers two reasons for thinking that collective intentional behavior is not the same as the summation of individual intentional behavior. The first is that the same type of bodily movements can on one occasion be a set of individual acts and on another occasion constitute a collective action. Searle illustrates this point by way of the following example:

Imagine that a group of people are sitting on the grass in various places in a park. Imagine that it suddenly starts to rain and they all get up and run to a common, centrally located shelter. Each person has the intention expressed by the sentence "I am running to the shelter". But for each person, we may suppose that his or her intentions is entirely independent of the intentions and behavior of others. In this case there is no collective behavior; there is just a sequence of individual acts that happens to converge on a common goal. Now imagine a case where a group of people in a park converge on a common point as a piece of collective behavior. Imagine that they are part of an outdoor ballet where the choreography calls for the entire corps de ballet to converge on a common point. We can imagine that the external bodily movements are indistinguishable in the two cases; the people running to the shelter make the same types of

bodily movements as the ballet dancers. Externally observed, the two cases are indistinguishable, but they are clearly internally different. (Searle, 1990: 402-403)

The problem then is to specify what this internal difference consists in. One important clue is that in the first situation the convergence on a common goal is a mere accident. The intention and behavior of each individual is entirely independent of the intentions and behavior of others. Moreover, as Searle remarks, this is so even if each person knows that the other people intend to run to the shelter and knows that the other people know that he or she intends to run to the shelter. Searle claims that in the second situation, by contrast, the individual "I intend"s are derivative from the "we intend"s. The second reason for thinking that collective intentional behavior is not the same as the summation of individual intentional behavior is that "often the derived form of an individual intention will have a different content from the collective intention from which it is derived" (Searle, 1990: 403). The example Searle gives is that of a football team trying to execute a pass play. No individual member of the team can have 'we are executing a pass play' as the entire content of his intention, for no one can execute a pass play by himself. Each player will must make a specific contribution to the overall goal.

The more controversial part of Searle's intuition is that collective intentions, what he calls we-intentions, cannot be analyzed into sets of I-intentions, even supplemented with beliefs, including mutual beliefs about the intentions of other members of a group. Notice that this claim is much stronger than the mere claim that collective intentional behavior is different from the summation of individual intentional behavior. The claim is that no analysis of collective intentions in terms of individual intentions is forthcoming whatever their combination is, whatever their contents are, and whatever other individual mental states they are supplemented with. Searle acknowledges that he cannot prove that no such analysis could ever succeed, but he claims that all reductive analyses that he has seen are subject to obvious counter-examples. The next section will consider what motivates Searle's introduction of we-intentions and whether the construal he offers of this notion should be accepted.

3. What are we-intentions?

According to Searle, reductive analyses fail to account for the cooperative and coordinated character of collective intentionality. More precisely, reductionist strategies, such as Tuomela and Miller's (1988), try to account for the cooperative dimension of collective actions in terms of mutual beliefs among members of a group. What Searle shows through counter-examples is that the existence of mutual beliefs is not sufficient to ensure cooperation. Thus, business school graduates who have been exposed to Adam Smith's theory of the hidden hand may come to believe that the best way for somebody to help humanity is by pursuing his own selfish interests. Each may form a separate intention to thus help humanity by pursuing his own selfish interests and not cooperating with anybody and they may all have mutual beliefs to the effect that each has such an intention. In such a case, despite all the businessmen having the same goal as well as mutual beliefs about their respective intentions, there is no cooperation and no collective action. What they lack is an intention to cooperate mutually. Mutual beliefs among members of a group do not ensure the presence of such an intention.

According to Searle, this cooperative dimension of collective actions can be captured only if it is accepted that the intentions attributable to the individuals that take part in collective actions are different in type from the intentions attributable to those same individuals when they engage in individual actions. The idea then is that to account for cooperation we have to introduce a specific type of mental states: we-intentions. What needs to be spelled out is the sense in which we-intentions are special and, relatedly, the sense in which they can be said to imply cooperation. *Prima facie*, three possibilities are conceivable.

The first is that what makes we-intentions special is features of their contents and thus that the dimension of cooperation is linked to specific features of these contents. The second is that what makes we-intentions special has to do with the type of entities they can be attributed to. Finally the third possibility is that rather than the contents or the possible bearers of we-intentions, it is the psychological mode itself — i.e., the fact that the psychological mode is that of we-intending instead of I-intending — that implies the notion of cooperation. Before we examine each possibility in turn, let me note that they are not mutually incompatible: we-intentions could in principle be special in all three ways.

The first possibility is that we-intentions are special in that their contents have specific features not shared by I-intentions. One could then claim that what is specific of the content of a we-intention is that I-intentions, related in a certain way, are embedded in it. To borrow Searle's example of two cooks, say Paul and Gilbert, preparing a hollandaise sauce together, the content of the we-intentions would be something like (that this collective intention causes Paul to have the I-intention to stir while Gilbert is pouring and Gilbert to have the I-intention to pour while Paul is stirring). According to this conception, what would make we-intentions a *sui generis* type of mental states are certain unique features of their contents, namely that further intentions fall in their scope, and what would capture the dimension of cooperation would be the way those embedded intentions are related. This is not the option Searle favors. Searle thinks such an analysis cannot be right for two reasons. First, because it would involve attributing two separate intentions to each agent, for instance in the case of Paul a we-intention to prepare the sauce

and an I-intention to stir. Second, because the we-intention of each agent would have to be an intention to make it the case that he have a singular intention. Although Searle is not explicit why he thinks that these consequences of this analysis are unacceptable, I gather that the reason why he rejects them is that they do not seem to accord with the phenomenology of the agents involved in a cooperative activity.

What he proposes instead is that the content of collective intentions is of a form already present in some complex cases of singular intentions. Namely, the content of the intention encompasses a by-means-of relation. The idea is that in the case of singular intention of, e.g., firing a gun by pulling the trigger, there is only one intention and one action, with the relation of the means-intention to the overall intention being only part-whole. Similarly, for Searle, in the case of collective actions, there is only one complex: the singular intentions of the agents are related to the collective intentions as means to ends and this relation of the singular intentions to the collective intention is simply part-whole. It is important to note that there is nothing in the by-means-of relation *per se* that implies cooperation. For instance, I can intend that we go to the police station by means of me dragging you, and clearly in such a case no cooperation need be involved. Thus, there is nothing in the analysis Searle offers of the form of the content of collective intentions that makes it necessary that the dimension of cooperation essential to collective intentions be reflected in their contents. Indeed, Searle insists that: "The real distinction between the singular and the collective case is in the type of the intention involved, not in the way that the elements in the conditions of satisfaction relate to each other" (1990: 412).

Let us now turn to the second possibility, namely that the specificity of collective intentions is related to the type of entities we-intentions can be attributed to. The idea here is that whereas I-intentions are typically attributed to individuals, there are further constraints an entity must meet for we-intentions to be attributable to it. This move can be made in two different ways. One is to admit the existence of some forms of primitive collective entities, such as group minds, collective consciousnesses, or Hegelian world spirits. Primitive here means that those entities are conceived as *sui generis* and not as constructions assembled from more elementary entities (in the way, say, that molecules are collections of atoms assembled in a certain way). Searle, quite reasonably I take it, sees the existence of such entities as widely implausible. He intends his account of collective intentionality to be consistent with methodological individualism, that is with the fact that society consists of nothing but individuals and that all the intentionality there is in the minds of individuals. Thus, it will not do to say that collective intentions form a specific type of states insofar as they are the privilege of primitive collective entities. But there is also an other move that can be made. Instead of saying that we-intentions are to be attributed to primitive collective entities, one could suggest that the proper bearers of we-intentions are individuals related in a certain way both among themselves and with their environment. Of course for such a strategy to have a chance of success, the relations that must obtain among individuals for them to qualify as bearers of we-intentions would have to be describable in a non-circular way. In other words, it should not be the case that the ability to enter into such relations presupposes collective intentionality. Whether a non-circular account of these intersubjective relations is feasible is something we shall consider in the next section. It should be noted however that this kind of move would be rejected by Searle. For Searle wants his account of collective intentionality to meet a further constraint besides methodological individualism. Searle is a dyed-in-the-wool internalist. Not only does he want all intentionality to be in individual minds or brains, he also maintains that all the intentionality an individual has could be had by this individual even if he or she were a brain in a vat. In other words, Searle also wants an account of collective intentionality to meet the following constraint: "It should be consistent with the fact that the structure of any individual's intentionality has to be independent of the fact of whether or not he is getting things right, whether or not he is radically mistaken about what is actually occurring" (1990: 406). The reason why adherence to this constraint is incompatible with pursuing the strategy just described should be rather obvious. For a relation to obtain, the relata must exist. But to say, as Searle does, that I could have all the collective intentionality I have if I were a brain in a vat, is to say that I could have all the collective intentionality I have even if there existed no other individuals for me to be related to.

So, the only possibility left open for Searle is the third one, namely, that what makes we-intentions special is the psychological mode itself, not the possible subjects of we-intentions nor their contents. Now, Searle also wants his view of collective intentionality as a primitive form of intentionality to satisfy the two constraints of individualism and internalism, the more so since the reason why he rejects some alternative proposals is that they flout either individualism or internalism. He maintains that these two constraints are in fact rather easily satisfied. In order to see that his conception of collective intentionality as primitive is consistent with individualism, it is enough, according to Searle, to "note that all the intentionality needed for collective behavior can be possessed by individual agents even though the intentionality in question makes reference to the collective" (1990: 407). And to see that it is consistent with internalism, it suffices to note that "collective intentionality in my head can make a purported reference to other members of a collective independently of the question whether or not there actually are such members" (1990: 407).

This analysis appears less than satisfactory for at least two reasons. First, Searle's contention that a single individual, even if he happens to be a brain in a vat, can have a collective intentionality seems difficult to understand. Searle himself acknowledges that his analysis has an uncomfortable feature. Namely:

[I]t allows for a form of mistake that is not simply a failure to achieve the conditions of satisfaction of an intentional state and is not simply a breakdown in the Background. It allows for the fact that I may be mistaken in taking it that the "we" in the "we intend" actually refers to a we, that is, it allows for the fact that my presupposition that my intentionality is collective may be mistaken in ways that go beyond the fact that I have a mistaken belief. [...] on my account, it turns out that I can not only be mistaken about how the world is but am even mistaken about what I am doing. If I am having an hallucination in supposing that someone else is helping me push the car, that I am only pushing as part of our pushing, then I am mistaken not only in my belief that there is somebody else there pushing as well but also about what it is that I am doing. I thought I was pushing as part of our pushing, but that is not in fact what I am doing. (1990: 408)

It seems to me that this admission of Searle's clearly contradicts his claim that his analysis of collective intentionality is consistent with internalism, i. e., consistent with the fact that all intentionality, whether collective or individual, could be had by a brain in a vat or a set of brains in vat. For what exactly is the special form of mistake Searle is referring to in the passage just quoted? Searle says that the hallucinating individual is mistaken about what he is in fact doing. How should we understand this claim? Searle (1983) identifies an action not just with a series of bodily movements, but with the bodily movements together with the intention in action that causes them.¹ Clearly it is not just the case that the hallucinating individual is mistaken about the kind of bodily movements he is performing, for this individual could be right that he is actually pushing the car and yet be mistaken in supposing that someone else is helping him. So in saying that the individual is mistaken about what he is in fact doing, Searle presumably means that he is mistaken about the kind of intention in action he has. Searle says that the individual may be mistaken in taking it that the "we" in the "we intend" actually refers to a we. But what could this mean other than that what the individual assumes is a collective intention in action is not actually a genuine one? What could this mean other than that the hallucinating individual is deluded in thinking that he has a we-intention. For Searle's analysis to be consistent with internalism it would have to be the case that a brain in vat can have genuine collective intentions, not just it can think it has. But if the brain in a vat is wrong in thinking that it has a collective intention, then clearly it doesn't have one. Therefore, it seems that, contrary to what Searle claims, his analysis is not consistent with internalism. An individual cannot properly be said to have a collective intention unless other individuals actually share his intention. A second problematic feature in Searle's analysis has to do with the dimension of intersubjectivity, a dimension that is glossed by Searle in terms of agreement and cooperation. Recall that, according to Searle, the businessmen pursuing their own selfish interests in order to help humanity would have constituted a case of collective intentionality if they had all got together on graduation day and agreed to so act. Recall also, that according to Searle, what the Tuomela-Miller analysis fails to account for is the notion of cooperation that is built into collective action. Well, does Searle's analysis fare any better? Nothing in the structure of the content of we-intentions as laid out by Searle seems to capture the notion of cooperation. As we have already noted, there is nothing in the by-means-of relation *per se* that implies cooperation. I can intend that we go to the police station by means of me dragging you, and clearly in such a case no cooperation need be involved. So how does cooperation enter the stage? According to Searle, in order to account for the cooperative character of we-intentions, we must appeal to Background capacities. What collective intentionality presupposes is "a Background sense of the other as candidate for cooperative agency; that is, it presupposes a sense of others as more than mere conscious agents, indeed as actual or potential members of a cooperative activity" (1990: 414). Now, Background capacities, according to Searle, are not themselves representational. Rather, they are a set of nonintentional or preintentional capacities that enable intentional states of function. In other words, they are biological or neurophysiological phenomena rather than intentional phenomena. As Fissette (1997) points out, by drawing a line between the realm of the intentional and the Background, and by considering that the dimension of cooperation is part of the Background, Searle acknowledges that he cannot account for it in intentional terms. From this it follows, that a theory of collective intentionality cannot *per se* provide a conceptual foundation for the social sciences. If cooperation is indeed an essential dimension of social phenomena, and cooperation is part of the Background, providing a conceptual foundation for the social sciences is ultimately a job for the biologist, not for the philosopher. To sum up, I disagree with Searle on two counts. First, to insist that an account of collective intentionality should meet the constraint of internalism seems to me close to an absurdity. It is one thing to claim that all the intentionality there is, including collective intentionality is in the head of individuals. It is another to insist that collective intentionality could be had by a single individual, not to mention a brain in vat. The first claim is at least plausible, the second seems to me unintelligible. Second, while there is no denying that all our cognitive capacities rest on an underlying basis of biological capacities, it

appears to me unduly hasty to sweep cooperation under the rug of Background presuppositions. Although a capacity for cooperation may depend in part on certain background biological capacities, I think it can also be construed in part in intentional terms. These two difficulties with Searle's account are not unrelated. It is, I think, Searle's adherence to internalism that prevents him to see how cooperation can be accounted for, at least in part, in intentional terms.

4. An alternative proposal

In this section, I will briefly present another proposal, made by Bratman, which I think fares better with regard to the elucidation of the main features of collective intentionality, even though it is not without its own problems. Bratman agrees with Searle that collective intentional behavior is not analyzable as just the summation of individual intentional behavior even supplemented by mutual beliefs or mutual knowledge. Yet contrary to Searle, he thinks that collective intentionality is not a primitive phenomenon and that a non-circular reductive analysis may be possible. Bratman (1992) identifies three features of shared cooperative activities that such an analysis would have to account for. The first feature is the mutual responsiveness of each participating agent to the intentions and actions of the others. The second is a commitment to a joint activity. The third is a commitment to mutual support by which each agent is committed to supporting the efforts of the other to play her role in the joint activity. None of these features is by itself sufficient to make an activity a shared cooperative activity, but, according to Bratman, taken together these three features are characteristic of shared cooperative activities.

Bratman is careful to ensure that his analysis is non-circular. In this regard the notion of a commitment to a joint activity may seem suspect. First, it may appear that cooperation is built into the notion of joint activity. Here, Bratman is careful to distinguish joint act types that are cooperatively neutral from those that are cooperatively loaded. For instance, we can go to New-York together without our activity being cooperative. He makes it clear that in his analysis the feature of commitment to a joint activity should be read in a cooperatively neutral way. Second, one may wonder whether it makes sense to appeal to my intention that we do something together, insofar as it may be thought that one can only intend one's own actions. Here, Bratman introduces a distinction between intending and attempting and defends a planning conception of intentions that emphasizes the role of future-directed intentions as elements of partial plans.² This conception of intentions allows him to be more liberal about what can be intended than about what can be attempted since references to things other than our own actions can function appropriately in our plans.

Since Bratman construes commitment to a joint activity in a cooperatively neutral way, this commitment does not suffice to ensure that the activity that follows is a shared cooperative activity or a collective action in Searle's sense. The originality of Bratman's analysis comes from the way in which he construes the two further features of mutual responsiveness and commitment to mutual support. These are analyzed in terms of meshing subplans and interdependent intentions. For an activity to be a shared cooperative activity, it must be the case that each agent intends that the group performs this joint action in accordance with subplans that mesh, where for subplans to mesh it is not necessary that there be full agreement in the agents' subplans, but merely that there be some way the action can be done that would involve the successful execution of those individual subplans. For instance, if John and Mary intend that they paint the house together, but Mary intends that they paint it red all over, and Peter that they paint it blue all over, then their respective subplans concerning the color of the paint don't mesh. If, by contrast, Peter intends that they paint the house blue all over but has no preference as to where they should buy the paint, whereas Mary does not care about the color but intends that they buy the paint at a particular store, then their respective subplans mesh. This meshing of subplans in turn implies that the intentions of the participating agents must be interlocking. Each agent should bring into the content of his intention the efficacy of the other participants' intentions as well as the efficacy of his own intentions. By thus requiring that the intentions of the participating agents be interlocking, Bratman moves away from the classical reductive analyses of collective action according to which the crucial link among the attitudes of the participating agents is simply cognitive. Mutual belief or mutual knowledge is not sufficient to ensure that intention is shared or collective; but neither is it necessary to posit collective intentionality as a primitive form of intentionality. What is crucial is the specific form of interdependence of the individual intentions. Finally, insofar as the intention of each participant includes his intending that the relevant intentions of the others be successfully executed, commitment to mutual support is also involved in shared cooperative activities. As Bratman phrases it, the intentions of the participants should be minimally cooperatively stable. In other words, there must be at least some circumstances in which each participant would be prepared to help the others do their part in the joint activity.

These conditions essential for shared cooperative activity are summarized by Bratman (1992: 338) in the following way:

Where J is a cooperatively neutral joint-act type, our J-ing is a shared cooperative activity only if:
(1) (a) (i) I intend that we J.

- (1) (a) (ii) I intend that we J in accordance with and because of meshing subplans of (1) (a) (i) and (1) (b) (i).
- (1) (b) (i) You intend that we J.
- (1) (b) (ii) You intend that we J in accordance with and because of meshing subplans of (1) (a) (i) and (1) (b) (i).
- (1) (c) The intentions in (1) (a) and (1) (b) are minimally cooperatively stable.
- (2) It is common knowledge between us that (1)

Here, conditions (1) (a) (i) and (1) (b) (1) are meant to capture the feature of commitment to a joint activity; Conditions (1) (a) (ii) and (1) (b) (ii) to capture the feature of mutual responsiveness, and condition (1) (c) the feature of commitment to mutual support.

Bratman's analysis appears to me more satisfactory than Searle's. It does not lay open to the criticism Searle addresses to traditional reductive analyses of collective intentionality since it does not maintain that the crucial link among the attitudes of agents involved in collective behavior is a purely cognitive link. It brings to the fore three essential features of shared intentions, that are not taken into account by traditional analyses. Finally, contrary to Searle, it tries to capture what is distinctive of shared intentions in terms of a special kind of interdependence of the individual intentions of the participants, rather than by postulating a mysterious form of primitive collective intentionality supposed to be in the head of individual agents. Earlier, I distinguished three different ways in which we-intentions could be conceived as special. According to the first, what makes we-intentions special are features of their contents, according to the second what makes them special has to do with the type of entities they can be attributed to, and, according to the third, what is special is the psychological mode itself. We have seen that the option favored by Searle was the third one. I think Bratman's position can be characterized as a mix between the first and the second option. According to Bratman, the contents of the intentions of individuals involved in collective action are special insofar as they make reference to the intentions of the other participants — each agent must have intentions in favor of the efficacy of the intentions of the others — and are reflexive as well — each agent must have intentions concerning the efficacy of their own intentions.³ But collective intentions seem also to be special insofar as they cannot only be attributed to individuals related in a certain way. Indeed, Bratman's choice of words provides a good indication. He does not speak of collective intentions, but rather of shared intentions. Trivially, for an intention to be shared, there must be at least two individuals to share it. Thus, on Bratman's view, the idea that a brain in a vat could have a collective intention is simply incoherent. A brain in a vat even if the content of his intention makes purported reference to the intentions of purported others does not have a collective intention. To speak of a collective intention, it is necessary that the intentions of each participant mesh with the intentions of the others, hence that there exist others.

Despite its illuminating character, Bratman's analysis is not totally unproblematic. The condition of mutual knowledge he proposes is notoriously difficult to satisfy and one may wonder whether such a strong condition is really necessary. In particular, if one supposes that mutual explicitation of intentions and plans requires resorting to verbal communication, it becomes doubtful whether an analysis such as Bratman's could really be useful given Searle's general project. For Searle, a capacity for collective behavior is a condition of possibility of the institution of language, and not vice-versa, even though of course, language makes possible new forms of collective behavior. If we think that the satisfaction of Bratman's mutual knowledge condition requires resorting to language, we cannot without circularity substitute his analysis of shared intention to Searle's while preserving Searle's general explanatory project.

A second drawback of Bratman's analysis, at least relative to Searle's overall project, is that the characterization he offers seems tailored to forms of shared cooperative activities that are already rather sophisticated. Bratman's analysis focuses on future-directed intentions, that involve rational deliberation and conscious planning, whereas Searle's analysis is at the level of intentions in action. One may therefore wonder whether his analysis still applies when one considers more elementary forms of collective behavior.

5. What cognitive capacities are needed for collective intentions and actions

One way to approach this question is to examine what kinds of cognitive capacities one must attribute to an agent for her to be able to be a participant in collective actions. By proceeding to such an examination, we may be able to do two things. First, we may be able to offer a more detailed articulation of the kind of capacities Searle locates in the Background, where these capacities may be characterized at least in part in psychological terms rather than directly in biological terms. Second, this may help us define a notion of mutual accessibility of intentions less demanding than Bratman's requirement of mutual knowledge. Obviously, it would be impossible to do justice in the space of a few paragraphs to the enormous literature already available. What I hope to do is simply to give some idea of the kinds of investigations potentially relevant to a characterization of the cognitive capacities that underlie collective intentionality.

I shall focus on two essential capacities and their interplay. The first is a capacity for planning. Collective action presupposes a capacity for coordinating one's own intentions and actions with those of the other participating agents. However, as Bratman's (1987) work shows, this capacity for planning and coordination is not required solely for collective action. Even in the case of individual actions, intrapersonal coordination is needed. In particular, individual intentions and actions are subject to consistency constraints. First, plans must be internally consistent: the intentions that are the constitutive elements of a plan must be mutually consistent, for it to be possible that the entire plan be successfully executed. Second, plans must also be 'externally consistent', that is consistent with the beliefs of the agents about the world she is in. Third, there is also a demand for means-end coherence: the means chosen must be adapted to the end pursued and, although plans are initially typically partial, they must be appropriately filled in as time goes by.

Bratman is mainly concerned with future-directed intentions and thus with forms of high-level planning. But planning and coordination take place at a number of levels, from the basic level of simple motor commands up to conscious deliberation about plans. For instance, work in the cognitive neuroscience of action shows that even actions as simple and seemingly automatic as eye saccades are controlled by sets of predictive mechanisms that allow for an internal simulation of eye movements in the absence of overt saccades and thus allow for a selection to be made among alternative actions (Berthoz, 1996). Of course the neural mechanisms that control eye-saccades are low-level mechanisms whose workings are not consciously accessible. Therefore, by Searle's standards the types of simulations they allow would not count as mental simulations: we are in the domain of the non-intentional, biological capacities that for Searle belong to the Background.

Yet, there are intermediate levels of planning between those that depend on these low-level neural mechanisms and the high-level capacities considered by Bratman. Of particular interest are those levels of planning through simulation that can give rise to conscious experience and thus to what is called motor imagery. According to researchers working on motor imagery (Decety & Ingvar, 1990; Decety, 1996, Jeannerod, 1994, 1997), the same neurophysiological mechanisms subservise both imagined and actually executed actions. There is thus a close relationship between motor imagery and what Searle (1983) calls the experience of acting — that which gives us a conscious access to the intentional content of our intentions in action. As emphasized by Jeannerod (1994), the experimental study of motor imagery may thus be a perspicuous way of investigating the content of intentions in action and of clarifying both what its elements are and how they are encoded⁴. Moreover, the study of motor imagery may shed light on a form of planning that is pitched at the level of intentions in action. Thus, the existence of a capacity for conscious motor simulation may be a key to understanding how some forms of intrapersonal as well as interpersonal coordination can be achieved. Of course this capacity will ultimately depend on more primitive capacities that are non-intentional and can only be characterized in purely neurophysiological terms, but it would be premature to claim that all the capacities needed for collective intentionality belong to the Background.

A second capacity that one must possess in order to participate in collective actions is a capacity to attribute mental states to oneself and others and to explain and predict behavior on that basis. How this capacity should be construed is currently the object of much debate. On the one hand, theory theorists maintain that in order to impute mental states to herself and others an individual must possess a body of knowledge about cognition and motivation that is theory-like. On the other hand, the simulationist approach contends that such a capacity can be construed as a capacity to imaginatively identify with someone else and to imagine the situation the person is in. One important difference between the two approaches is that on the theory-theory view engagement in folk-psychological practice requires the possession of psychological concepts such as the concept of belief, whereas on the simulationist approach, such mastery is not required. On the latter view, it is enough, for instance, that by imaginatively identifying with someone else, one imaginatively believes (desires, intends, ...) this or that. There are at present various versions of each approach as well as hybrid variants that take the two approaches to be complementary rather than rival.⁵

Certain recent data in the neurophysiology of action seem to provide evidence in favor of the simulationist approach. A new class of visuomotor neurons has been recently discovered in the monkey's premotor cortex (Rizzolatti *et al.* 1988; Di Pellegrino *et al.* 1992). These so-called mirror neurons respond both when a particular action is performed by the recorded monkey and when the same action, performed by another individual, is observed. Mirror neurons thus appear to form a cortical system matching observation and execution of goal-directed motor-action. These findings have been extended to human subjects in a recent studies of PET studies (Decety *et al.*, 1994, 1997; Grafton *et al.*, 1996, Rizzolatti *et al.*, 1996, Stephan *et al.*, 1995). These studies provide evidence for the existence of a cortical network common to conditions where subjects are intending actions, preparing for execution, mentally simulating actions and observing actions performed by other individuals. As suggested by Gallese and Goldman (1998), one possible function of these structures may be to enable an organism to

detect certain mental states of observed conspecifics. This function may be part of, or a precursor to, a more general mind-reading ability relying on a capacity to adopt a simulation routine.

A general discussion of the debate between theory theory and simulation theory falls beyond the scope of this paper and my purpose is certainly not to try to adjudicate between the two approaches. Rather, my point is that the existence of such philosophical and psychological theorizing shows that rather than directly appealing to a biologically primitive sense of the other person as a candidate for shared intentionality, one can usefully investigate the nature of the psychological mechanisms that enable human beings to attribute mental states to others and to see things from their perspective.

It should be noted that neither of the two capacities just discussed — a capacity for planning and coordination and a capacity for mind-reading — is specifically tied to collective action. Both capacities are already involved in individual behavior. Yet, for collective intentionality to be possible, it must be the case that these two capacities are integrated and thus work together. In particular, this integration is necessary for the interlocking of intentions which, according to Bratman, is a central characteristic of shared intentions. In other words, for shared cooperative activity to be possible, it is not enough that each individual be capable of attributing to the other participants intentions similar to his own. It is also necessary that (1) each agent be capable of imagining both the third-person information available to the other agents in the situation considered and their first-person orientation toward this situation, (2) that each agent be capable of imagining the third-person information and the first-person orientation that the others attribute to him, and (3) that each agent be capable of repeatedly switching perspective in order to reach interpersonal coordination⁶. Presumably, the integration of planning and mind-reading capacities requires a fair amount of cognitive resources, including a capacious working memory and a relatively sophisticated level of executive function. Yet, once again those working memory and executive function capacities are not specifically tied to collective intentionality, they are also involved in a number of cognitive activities, such as problem-solving, that are not essentially linked to collective behavior. What this brief sketch of the kind of capacities involved in collective intentions and actions is meant to suggest is that, in order to account for the capacity to engage in collective behavior, one does not need to appeal (at least in the first instance) to specific Background capacities. An explanation can be couched in terms of the psychological endowment of human beings, where the psychological capacities appealed are not specifically dedicated to the production of collective behavior but fulfill a number of roles in the cognitive life of human beings. This discussion of the capacities underlying collective intentionality and action may also help see how to replace Bratman's requirement of mutual knowledge with a weaker requirement. Rather than of mutual knowledge, which seems to require explicit communication, I propose that we speak of mutual presumption, namely the presumption that the other agents are sufficiently cognitively similar to us that their attitudes and intentions can be successfully simulated or inferred. Because such a presumption can be mistaken, our intentions regarding collective actions are fallible, in other words, they are themselves presumptive. It follows that the actual existence of shared intentions is not something that can be established with certainty prior to the action itself. It is the satisfaction of the mutual expectations of the agents, the support given when problems arise that warrant the reality of presumptive shared intentions.

6. Concluding remarks

According to the proposal put forward here, it is neither the case that collective intentionality is reducible to individual intentions in the way classical analyses suggest, nor that it is a primitive form of intentionality backed up by specific Background capacities. Rather the specificity of collective intentionality lies in the interlocking character of the system of intentions involved. Thus, there is no need to posit a *sui generis* kind of psychological attitude, a we-intention, to account for collective intentionality. The common breed of intentions will do the trick, but their contents will be more complex than they are in the case of individual actions. What about Searle's requirements that an account of collective intentionality must be consistent with both individualism and internalism? The proposal advanced here is certainly consistent with individualism. All the intentionality there is in the head of individuals. The difference with Searle's account is to be found in the requirement that for the intentions of the individuals together to constitute collective intentions, they must in fact be shared and not just be taken to be shared.

The issue of internalism is more complex. Let us go back first to what Searle himself sees as an uncomfortable feature of his analysis, namely that it allows for a form of mistake that is not simply a failure to achieve the conditions of satisfaction of an intention but a more radical mistake concerning what type of state one is in. In other words, the hallucinating individual or the brain in a vat who thinks he is we-intending to do something is not only mistaken in his beliefs about the world, he is also mistaken in taking it that his intention is a we-intention. By contrast, the alternative analysis offered here does not allow for such a mistake, since it avoids postulating that the intentions possessed by agents involved in collective behavior are different in kind from individual intentions.

To answer the question whether, on the proposed account, a brain in a vat could have a collective intention, one must first draw a distinction. On this account, an intention of an individual can properly be called a collective intention only insofar as it is part of an interlocking system of intentions. Obviously, if collective intentions are defined in this way a brain in a vat cannot be said to have collective intentionality. But the question whether a brain in a vat can have collective intentionality can also be taken in a weaker sense. In this weaker sense, it becomes the question whether a brain in a vat could have an intention of a type such that it could, if circumstances are right, constitute an element in an interlocking system of intentions. Here, the answer may well be positive. A would-be collective intention — that is, an intention fit to become an element in an interlocking system of intentions — is an intention the content of which makes reference in a certain way to the intentions of other members of a group. Provided one does not already have general objections to internalism, this case does not create problems of its own for internalism and thus supplies no new grounds to reject it.

References

- Berthoz, A. (1996). Neural Basis of Decision in Perception and in the Control of Movement. In A. R. Damasio, H. Damasio, & Y. Christen (eds), *Neurobiology of Decision-Making*, Berlin: Springer-Verlag, pp. 83-100.
- Bratman, M. E. (1987). *Intentions, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.
- Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review*, 101, 2, 327-41.
- Carruthers, P. & Smith, P. K., eds (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Davis, M. & Stone T., eds (1995a) *Folk Psychology*. Oxford: Blackwell.
- Davis, M. & Stone T., eds (1995b) *Mental Simulation*. Oxford: Blackwell.
- Decety, J. (1996). The neurophysiological basis of motor imagery. *Behavioural Brain Research*, 77, 45-52.
- Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., & Fazio, F. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain*, 120, 1763-1777.
- Decety, J. & Ingvar, D. H. (1990). Brain structures participating in mental simulation of motor behaviour: a neuropsychological interpretation. *Acta Physiologica Scandinavica*, vol. 73, 13-34.
- Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., & Fazio, F. 1994. Mapping motor representations with PET. *Nature*, 371, 600-602.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. 1992: Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91, 176-80.
- Gallese V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, vol. 2, 12, 493-501
- Grafton, S. T., Arbib, M. A., Fadiga, L. & Rizzolatti, G. 1996: Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Experimental Brain Research*, 112, 103-111.
- Jeannerod, M. (1994). The representing brain: neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187-246.
- Jeannerod, M. (1997). *The cognitive Neuroscience of Action*. Oxford: Blackwell.
- Fisette, D. (1997). Intentionnalité collective, rationalité et action. In J.-P. Dupuy & P. Livet (eds), *Les limites de la Rationalité*. Paris: La découverte.
- Pacherie, E. (1997a). Motor images, self-awareness, and autism. In J. Russell (ed.), *Autism as an Executive Disorder*. Oxford: Oxford University Press, p. 215-255.
- Pacherie, E. (1997b). Troubles de l'agentivité et troubles de la conscience de soi: quelques hypothèses sur leurs liens dans l'autisme. In J.-L. Petit (ed.), *Les neurosciences et la philosophie de l'action*. Paris: Vrin, pp. 363-385.
- Pacherie, E. (1998a). "Représentations motrices, imitation et théorie de l'esprit", in H. Grivois & J. Proust (eds), *Subjectivité et Conscience d'Agir: Approches Cognitives et Cliniques de la Psychose*, Paris, PUF, pp. 207-248.
- Pacherie, E. (1998b). The content of intentions. *Rapports et Documents du CREA*, N° 9724, Février 1998. To appear in *Mind and Language*. Rizzolatti, G., Carmada, R., Gentilucci, M. Luppino, G & Matelli, M. 1988: Functional organization of area 6 in the macaque monkey. II Area F5 and the control of distal movements. *Experimental Brain Research*, 71, 491-507.
- Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D. & Fazio, F. 1996b: Localization of grasp representations in humans by PET. 1. Observation versus execution. *Experimental Brain Research*, 111, 246-252.
- Searle, J. R. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. R. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, & M. E. Pollack (eds), *Intentions in Communication*. Cambridge, Mass.: MIT Press.
- Searle, J. R. (1995). *The Construction of Social Reality*. New-York: The Free Press.
- Stephan, K. M., Fink, G. R., Passingham, R. E., Silbersweig, D., Ceballos-Baumann, A. O., Frith, C. D., & Frackowiak, R. S. J. 1995: Functional anatomy of the mental representation of upper extremity movements in healthy subjects. *Journal of Neurophysiology*, 73, 373-386.

NOTES

¹ Searle (1983) proposes a distinction between two types of intentions, what he calls intentions-in-action and prior intentions. In his terminology, a 'prior intention' corresponds to the initial representation of the goal of the action prior to the initiation of the action, whereas an intention in action is the proximal cause of the physiological chain leading to overt behavior. Prior intentions and intentions in action are said to

differ in their contents: Searle claims that whereas the content of intentions in action presents physical movements, the content of prior intentions represents whole actions, that is, not just a physical movement, but the causal sequence consisting of the intention in action causing the physical movement. Moreover, Searle points out that the content of an intention of action is much more determinate than the content of a prior intention, meaning that my intention in action to raise my arm, for instance, will include not only that my arm goes up, but that it goes up in a certain way, at a certain speed, etc. According to Searle, another important difference between prior intentions and intentions in action that not all intentional actions have prior intentions in actions but all intentional actions have intentions in action, where an intention in action does not simply trigger bodily movements but plays a continuing causal role in shaping them, guiding and monitoring them until completion.

² Bratman's (1987) theory of intentions emphasizes the role of intentions in planning and reasoning. Intentions are not only terminators of practical reasoning, they are also prompters of practical reasoning. This is what Bratman calls the reasoning-centered dimension of the commitment to action characteristic of intentions. Intentions play an important role as inputs into further practical reasoning, as, e. g., means-ends reasoning, reasoning from more general to more specific intentions, or reasoning as to which other intentions are consistent with a given intention. They are thus "typically elements in larger plans, plans which facilitate coordination both socially and within our own lives, plans which enable prior deliberation to shape later conduct" (1987: 28). Bratman distinguishes between future-directed intentions —concerned with future courses of action — and present-directed intentions — intentions to act in a certain way beginning now. Obviously, the reasoning-centered dimension of commitment is characteristic of future-directed intentions. Whatever practical reasoning the agents engages in, it will have occurred prior to the time of action. Future-directed intentions are thus central in his theory.

³ It should be noted that whereas for Searle all intentions, whether individual or collective involve a kind of self-referentiality, Bratman thinks that self-referentiality is a feature specific of collective intentional actions.

⁴ For much more detailed discussions on how work in the cognitive neuroscience of action can shed light on issues in the philosophy of action, see Pacherie, 1997a, 1997b, 1998a, 1998b.

⁵ On this debate, see for instance, Davies and Stone, 1995a, 1995b or Carruthers and Smith, 1996.

⁶ It may well be objected that these requirements are too strong. Indeed, it is doubtful whether certain social animals capable of collective actions — a pack of wolves performing collective hunting, for instance — are endowed with such sophisticated cognitive capacities. It is also doubtful whether each time human beings engage in collective action, those capacities are at play. Here, it may be useful to draw a distinction between the kinds of (individual or collective) actions that are part of the repertoire of a given species and are to some degree preprogrammed, and those that need to be learned. Another useful distinction is between learning in the sense of assembling a new action program and learning in the sense of adjusting parameters in a preexisting action program. Presumably, human beings learn how to drive a car in the first sense, but they learn how to walk only in the second sense. Now, it may be suggested that collective hunting in wolves is part of the innate repertoire of the species. If such is the case, it is certainly superfluous to attribute to them the kind of imaginative and perspective-taking capacities just listed, it is probably enough that they be able to exploit certain environmental and behavioral cues that are associated with given steps in the action program. Even given a prepackaged action program, wolves may have to learn how to hunt collectively, but then only in the second sense of learning how to adjust certain parameters. By contrast, it seems to me that learning novel types of collective actions requires the integration of planning and mind-reading capacities, even if, once those new forms of collective actions are well-practiced, the exploitation of behavioral cues may suffice for successful performance. In any case, if collective intentionality is to be used as a conceptual tool for explaining social and institutional facts, we should surely focus on the forms of collective intentionality that can give rise to novel types of collective actions.