

On Collective Intentions*

Kepa Korta

In those early amorphous years when memory had only just begun Esthappen and Rahel thought of themselves together as Me, and separately, individually, as We or Us.

Arundhati Roy: *The God of Small Things*, p. 2.

Unlike Esthappen and Rahel, I usually think of myself as Me, and I think as We or Us when thinking of myself together with other selves. I believe I am not a special case: most people (except, perhaps, real or fictional twins like Esthappen and Rahel) think of themselves in a similar way. When talking about Us, though, 'We' can be used in different occasions with different meanings. Leaving aside majestic uses and the like, sometimes 'We' merely refers to a set, a collection of people where 'I' is included. In this case, 'we' is nothing over and above the aggregation of individuals, an abbreviated manner of referring to her, you and me. But, sometimes we mean something stronger by 'We'. In this stronger sense, We is not yet a synonym for her, you and me; we somehow have been able to constitute together a new 'entity' called 'We'.

In this chapter, we will not discuss these different uses or meanings of the first-person plural pronoun. We will rather study the case of collective action; the sort of action whose agent is a 'we' in the second, strong sense; the sort of action that seems not to be explainable solely in terms of the actions of the individuals constituting such a 'we'. In fact, there is some consensus among philosophers on the view that there is such a thing as collective action, which is not just the sum of individual actions. However, there is no such a consensus on which its correct characterization is. Actions like singing a duet, dancing a tango, painting a house together, lifting the piano upstairs... are being intensely analyzed by philosophers like Raimo Tuomela, Margaret Gilbert, Michael Bratman and John Searle. The crucial concept is the concept of (*joint, shared, group, team, social, we-*) *collective* intention. The discussion of their proposals will be the main theme of this chapter.

* Earlier and substantially different versions of this paper were presented at the Sixth International Colloquium on Cognitive Science (ICCS-99) in Donostia, May 1999, and at the Seminar on Communication in Torino, October 1999. I wish to thank the audiences in both places for their helpful comments. This work was partially supported by a research grant of the University of the Basque Country (UPV I09.I09-HA010/99) and a research grant of the Spanish Ministry of Education (MEC HI1998-0051) for a joint project with the Centro di Scienza Cognitiva of the University of Torino. The major part of this version was completed during a stay at the Engineering and Science Lab in Harvard University. I wish to thank Barbara Grosz for making it possible. Thanks also to Jesus M. Larrazabal for his helpful comments on earlier drafts of the paper.

In the first section we will state as clear as possible the problem of collective action: how to reconcile the acceptance of the existence of collective action with the rejection of the existence of any primitive collective agent or mind. The second section will briefly stress the difference between collective action proper and a sum of singular actions. It is in the third section where we begin the discussion on the proposals mentioned above, with Gilbert's proposal and the risk to imply the existence of a group mind. Then, in the fourth section, which is the longest one, we will discuss Tuomela's, Bratman's and Searle's proposals, arguing for a primitive concept of collective intention. In the fifth section we will explain the place that this concept occupies in the explanation of collective action and agency. The sixth one will be devoted to the consideration of one important objection to any concept of collective intention. And the final section will be focused on some basic properties of collective intentions.

1. Collective action: The problem

Let us assume from the very beginning, together with all the authors considered here, that there is such a thing as singular action. Each person can act alone in accordance with her own beliefs, desires and intentions, without taking into account, in principle, any other person's beliefs, desires and intentions. Of course, people usually share the physical environment so that they must coordinate their actions with the actions of other people: think in our behavior as drivers in the city traffic. But this does not make our actions collective; they are just an aggregate of singular actions, be they "social" or not in the Weberian sense. On the other hand, when we see a couple dancing a tango, two singers in a duet, a group of people painting a house together we speak naturally of collective action and collective agency,¹ meaning something else than a mere sum of individual actions. At an intuitive level it seems evident that

"There really is such a thing as collective intentional behavior that is not the same as the summation of individual intentional behavior." Searle (1990), p. 402.

But philosophically this intuition turns out problematic. Any talk about the existence of something (action, mental state, agent) collective triggers, for some people, an immediate signal of danger. They surely think that the acceptance of anything collective necessarily undermines the so-called thesis of methodological individualism, which states that

¹ The terminology used by the different authors does not help clarifying the exact terms of the discussion. In the literature we find a variety of labels such as *collective*, *joint*, *shared*, *group*, *team*, *social*, *we-*, *common*, *mutual* for actions, intentions and beliefs, sometimes used as synonyms, sometimes in fancy combinations as *shared joint intention*, just for an instance. I will use *singular* versus *collective* actions, intentions, and beliefs. I prefer *singular* to *individual* or *personal*, because these when opposed to *collective* suggest that collective actions and attitudes are not actions and attitudes of the individuals, of the persons. In fact, *collective* is not without problems, since it seems to suggest that they are not attitudes of individual minds but of a Hegelian World Spirit, some supraindividual mind. The problem is precisely to explain collective action and agency without implying the existence of such enigmatic "supermind."

“All social phenomena ... should always be understood as resulting from the [mental states] and actions of individuals, and we should never be satisfied by an explanation in terms of so-called ‘collectives’.” Popper (1969), p. 98.

This is an ontological constraint. According to it, all social or collective behavior must be explained in terms of the mental states and actions of the individuals; in other words, any account of collective action should not involve in the ontology any other entity than individuals. Collective mental states ascribed to Hegelian collective minds are not to be accepted. Collective actions by mysterious superagents are not satisfactory explanations. All right. But accepting methodological individualism does not mean denying the existence of collective action. It only places a constraint, an important constraint, to our account of collective action. In fact, the authors that we will discuss here (Gilbert, Velleman, Searle, Tuomela, Bratman) all seem to accept this constraint. First, they accept the existence of collective action—this is what they want to explain. Second, they think that collective action is not just the mere aggregate of the actions of the individuals belonging to the relevant set of agents, not even when the agents share the same goal. But, third, they believe that, ontologically, a collective agent can be nothing over and above the individual agents that constitute it.² This means that the account of the existence of collective actions and agents should not be based on any postulation of collectives as ontological primitives. The only ontological primitives accepted are individuals. Quoting Searle again:

“[Collective intentionality] must be consistent with the fact that society consists of nothing but individuals. Since society consists entirely of individuals, there cannot be a group mind or group consciousness. All consciousness is in individual minds, in individual brains.” Searle (1990), p. 406.

But is it possible to make justice to our intuitions on collective action without violating this constraint? The proposals considered here try to show that it is.

2. Collective action versus aggregate actions

Being performed by more than one agent is, of course, a necessary condition for collective action. But it is not sufficient. As we said earlier, there can be a set of agents acting at the same time, even coordinately, without acting collectively. But what is more important, the same sort of bodily movements by a set of agents can count, in one case as a collective action, and as a mere aggregation of singular actions in another. Take a real example:³

² We are not addressing the case of structured social groups and institutions. Anyhow, the explanation of their existence is necessarily based on the explanation of the simplest cases of collective agency, which is the aim of our discussion. See Searle (1995), Tuomela (1995) or Gilbert (1989).

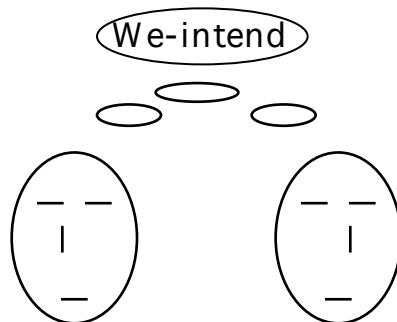
³ Bilbao, summer 1993.

In a holiday park where some hundreds of people are consuming drinks at the taverns especially opened for the occasion, a policeman appears in plain clothes and, recognized by some young guys, is beaten by them, and consequently seriously injured. In the action of beating the policeman each guy has the intention to beat him, but we can suppose, following the results of the judicial investigations well-known by now, that each guy's intention is in the relevant sense independent of the intentions and actions of the others. That was one of the main points to be established by the judge. He thought that in this case it was no collective action but a sequence of individual actions that happened to converge on a common goal.

So, the identification of the actions performed by a set of people either as a collective action proper or as aggregate singular actions cannot be determined solely by the bodily movements of the agents. We must consider the intentions of the agents. However, that an action is collective seems not to be fully explained by appealing to the singular actions and intentions of the agents involved, even when these actions have a common goal. We need a new type of intention, which we call collective intention. But what is this?

3. Collective intention and group minds

We want to characterize what a collective action is, i.e. in which consists an action performed by a collective agent. By analogy with the singular case, we can do it by ascribing to the agents a particular mental state called, *collective* (Searle), *shared* (Bratman, Gilbert) or *we-* (Tuomela) intention. All these terms may suggest, however, that it is not an intention of an individual, that it is not an intention of me or an intention of you, but an intention of us: something that, borrowing some graphical portraits from Searle (1995), might be represented like this:



This sort of view, to my knowledge, is explicitly defended, in the recent philosophical literature on the matter, only by David Velleman:

“... distinct intentions held by different people can add to a single token of intention, jointly held.” Velleman (1997), p. 31.

I think that “a single token of intention, jointly held” is the best description of the graphic above. Velleman wants to show how two or more individuals can literally share an intention. This aim immediately raises serious questions like where is this collective intention instantiated? In whose mind/brain? Not in mine, not in yours, but in ours? What does it mean? In answering these questions Velleman’s proposal is—as, I think, any other along the same lines should be—irremediably led to the following dilemma: either (1) the term “intention” is no longer used for a mental state but for some other thing that can be literally shared, or (2) methodological individualism is abandoned and there is room for collective minds and consciousness. Velleman is ready to admit both horns of the dilemma. Concerning the first,

“I am not sure that intention is essentially mental. There are of course mental intentions, but perhaps there can also be oral or written intentions—just as there are not only mental but also oral or written assertions.” Velleman (1997), p. 37.

Concerning the second,

“One may want to insist on intention’s being a mental state, of course. But then I would be inclined to say that the existence of collective minds remains an open question.” Velleman (1997), p. 38.

He then proposes to suspend judgment on both questions. But he is surely asking too much.⁴ What we want is an account of collective action that, first and obviously, does not deny the existence of collective action, second, keeps the concept of intention as referring to a mental state within the mental cause theory of action and, third, takes methodological individualism as a serious constraint. Abandoning any of these principles seems just question begging.

Tuomela, Searle, Bratman, and Gilbert all say that they want their proposals to be consistent with methodological individualism. However, it is not always clear that, in fact, they are. In particular, Margaret Gilbert, with her account in terms of “plural subjects,” seems often to imply the existence of a supermind. When she talks about shared intention, it is easy to understand that she is defending something similar to Velleman’s concept of collective intention as a “a single token of intention, jointly held”:

“...it seems that there could be a shared intention to do such-and-such though none of the participants **personally** intend to conform their behavior to the shared intention” Gilbert (1997), p. 68.⁵

or, some lines below

⁴ See also Stoutland (1997), who argues for *social* attitudes, which are not mental states, ascribable to *social* groups, without arguing for a supraindividual mind. The price to be paid is to abandon the causal theory of action.

⁵ My emphasis.

“... it is apparently possible in principle that the corresponding personal intentions be lacking when a shared intention is present.” Gilbert (1997), p. 68-69.

It is worth trying to clarify what Gilbert is stating. I think we can read her words in two different ways. On the one hand, notice that she says **personal** intention, not **individual**. If we understand by “personal intention” just “singular intention” in the sense proposed earlier, i.e., for the classical notion of intention considered traditionally in action theory, her claim would amount to saying that it is not necessary to have an intention of that kind once there is a shared or collective intention. If, according to her account, collective intentions were still mental states as individual as singular intentions, I would not pose any objection to those claims, and wait for her positive account of shared intentions. This interpretation seems to be consistent with her following words:

“Does it mean that a shared intention is not a function of the mental states of the participants? **No**, only that the relevant mental states are not personal intentions of the participants in favor of the corresponding behavior. Does it mean that shared intentions cannot motivate? **No**, only that insofar as they do motivate, their motivational force does not derive from the motivational force of corresponding personal intentions. Does it mean that a sensitive understanding of the nature of *intention* is not relevant to an understanding of shared intention? **Not necessarily**.” Gilbert (1997), p. 69.⁶

But these negative answers, on the other hand, do not sufficiently clarify her positions. Our doubts would have been removed had she answered “Yes, a shared intention is a (function of the) mental state(s) of the participants. Yes, shared intentions do motivate. Yes, a sensitive understanding of the nature of intention is relevant to an understanding of share intention.”

But, as it stands, Gilbert’s proposal can be taken to claim that shared intentions are not intentions of the individual minds, thus, implying either the existence of supraindividual minds or that shared intentions are not mental states.⁷ And, as said before, we want a theory which does not imply the existence of any group mind and which takes collective intentions as mental states.

4. Collective intention: can it be reduced or is it a primitive concept?

⁶ Emphasis added.

⁷ This is reinforced by the examples she uses: On a walk, Tina and Lena have a shared intention to turn back in half an hour, but Lena intends to act contrary to the shared intention. (Gilbert (1997), p. 68). This is even stronger that the compatibility of shared intentions with lack of the corresponding personal (singular) intentions. According to Gilbert, here Lena can have simultaneously the shared intention to go on a walk with Tina and the personal (singular) intention not to go. As far as I know, no theory of intention and action makes room for contradictory intentions. Velleman also construes Gilbert’s account of plural subjects as sounding “perilously close to talk of group minds or superagents.” Velleman (1997), p. 30.

Our account of collective action should explain how its existence is possible as different from the mere summation of singular actions, whereas we stick to methodological individualism, that is, to the ontological assumption that there are no primitive collectives, groups or ‘we’s. So the conceptual analysis goes in the following direction:

1. Collective intentions



2. Collective action



3. Collective agent, group, team, we...

(‘ $X \Rightarrow Y$ ’ is construed as ‘X (or its analysis) plays an essential role in the analysis of Y but not vice versa.’)

It is worth insisting that we assume that ontologically there should be no difference between a collective agent, a group or *we* and a mere set of agents. *We* are just you and I. But we are explaining the difference between, in the example above,

(a) ‘We (all) hit the policeman,’ that can be true even if you and me, each separately, hit him, with different singular intentions, and

(b) ‘We hit the policeman (together),’ that is not true in the case above, because *we* hit the policeman collectively. I hit the policeman and you hit the policeman but only as long as you and I are the constituents of this collective ‘we’.

The intuitive difference between these two cases seems evident.⁸ How do we explain it? Since we cannot appeal to any ontological difference between the ‘we’s in (a) and (b), we have to explain it recurring to the notion of ‘collective action.’ So a collective agent is just the term we use to refer to the (system of) agents of a collective action.⁹ For clarifying the notion of collective agent, thus, we have to analyze the notion of collective action. It is important to note that now it will not do to define collective action as the action whose agent is collective.¹⁰ That would make the analysis viciously circular. We have to define collective action recurring to the mental states of the agents that participate in it. This sort of mental states are what we call ‘collective intentions’, not because they are a mental state of a collective agent—this terminology is, of course, misleading, but we lack a better one—but because they define the actions of the agents as a collective action and, through this, the system of agents as a collective agent. We should keep in mind this direction of the conceptual analysis for approaching the main topic of the

⁸ That can be seen clearer, when we think of the responsibility of each agent in each case.

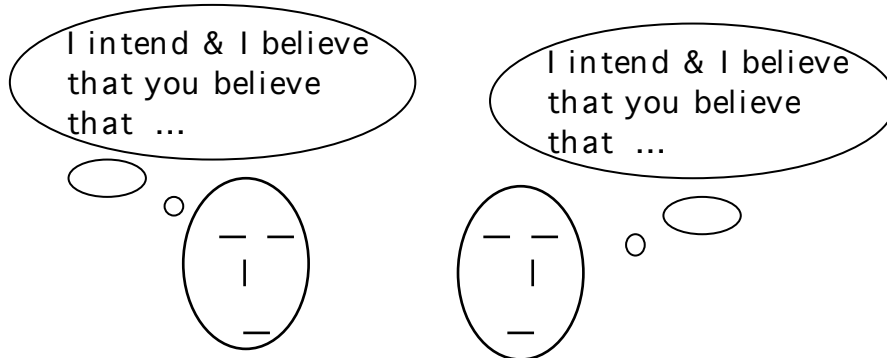
⁹ We are restricting our discussion to collective action and intentions. As far as I know, initially this was particularly the subject of discussion. It is true that Searle speaks about ‘collective intentionality’, which would include collective beliefs, goals, wishes, etc., but he lacks a detailed account of these notions. On the other hand, Gilbert (1987) and particularly Tuomela (1995, Balzer and Tuomela (1997), Miller and Tuomela (1999)) extend their analysis to collective beliefs, goals and attitudes in general. Concerning belief, it would be interesting to analyze what are the differences between collective belief, on the one hand, and the much better known concept of mutual belief, on the other.

¹⁰ As proposed, for instance, by Hobbs (1990).

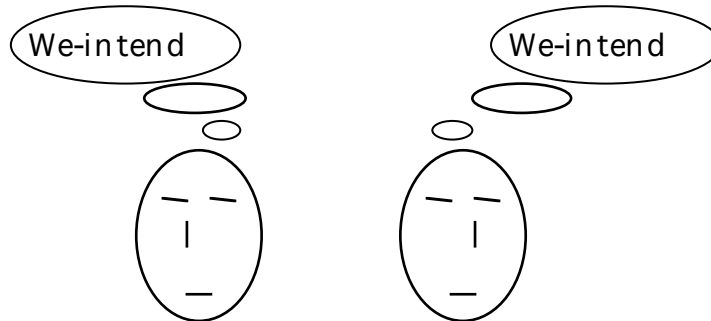
present section, namely, the discussion on the primitive or reductive analysis of collective intentions.

Once left aside those proposals that involve the existence of supraindividual minds, we can classify, following Searle (1990, 1995), the different explanations we are considering as belonging to one of these two groups:

Group A. Those who try to define collective intentions in terms of singular intentions and mutual beliefs.



Group B. Those who claim that collective intention is a primitive concept.



Of course, all the authors do not wholly agree on their ascriptions to one group or the other. According to Searle, the only author in B is himself, while Bratman and Tuomela (and Miller) are in Group A.

Bratman considers his own proposal “individualistic in spirit,” thus including himself in group A, while in B he includes Searle and Gilbert:

“In “Collective Intentions and Actions” John Searle argues that “collective intentional behavior is a primitive phenomenon” (401) and that we should eschew “a reductive analysis of collective intentionality” (406). In *On*

Social Facts Margaret Gilbert focuses on “plural subject concepts” (...) In contrast with both Searle and Gilbert, I have argued that a useful reduction may be possible here.” Bratman (1992), p. 341, fn. 25.

On the other hand, though sometimes his definitions seem to suggest the opposite, Tuomela has repeatedly claim that his proposal is not a reductive attempt:

“we therefore need a concept of we-intention which is not reducible to mere personal I-intentions”, Tuomela and Miller (1988).

“Searle incorrectly assumes in his paper that (conceptual) reduction of we-intentions to personal intentions and mutual beliefs was attempted in our paper. (...) Searle (1990) adopts a view of collective intentions which is rather close to my view.” Tuomela (1995), pp. 427-8, fn. 6.

So, it seems that, after all, Bratman is alone in group A.¹¹ But, is collective intention reducible to (definable in terms of) **singular** intentions? Searle (1990) thought that Tuomela and Miller’s analysis of *we-intentions* was just an affirmative answer to this question. Their analysis is roughly the following one:

The agent A, who is a member of the collective G, has the collective intention (joint or we-intention, as they call it) to do X if and only if:

- i. A intends to do her/his part of X;
- ii. A believes that the opportunities of success of X obtain; especially that a sufficient number of the remaining members of the group (at least probably) will do their parts of X; and
- iii. A believes that the members of the group mutually believe that the conditions expressed in (ii) obtain.

The fact they use the phrase ‘if and only if’ and A’s ‘collective intention’ on the left hand-side and simply ‘intends’ on the right hand-side is what makes their analysis appear to be a reductive attempt. If we analyze this as a reductive definition, so that A’s intention (i) is viewed as a mere singular intention the definition cannot differentiate between a collective action proper and a mere sum of singular actions. Take a simple version of the example above, with only two people, you and I, hitting the policeman. In this case, we can have

X= we hit the policeman

¹¹ Nevertheless, his point of departure seems to be the same as Searle’s, i.e., the intuition that collective behavior is not just the sum of individual behavior, on the one hand, and methodological individualism, on the other:

“On the one hand, it is clearly not enough for a shared intention to paint the house together that each intends to paint the house... On the other hand, a shared intention is not an attitude in the mind of some superagent consisting literally of some fusion of the two agents. There is no single mind which is the fusion of your mind and mine.” Bratman (1993), p. 98.

my part of X= I hit the policeman

I can have my singular intention to hit the policeman, thus satisfying (i); I can believe that you will hit the policeman, so that X will obtain, thus satisfying (ii); I can believe that this is mutually believed between us, satisfying (iii), and still we can be in case (a), a summation of singular intentions plus mutual beliefs, and not in case (b), a proper collective action of *us* (in the relevant collective sense) hitting the policeman together.

So, if understood as a reductive attempt, Tuomela and Miller's analysis fails. That is what Searle (1990) showed. But, as already remarked, we should not take it as a reductive definition. We should not interpret intention (i) as a singular intention: when making reference to 'his part of X' it is implicitly involving *our* doing it *together*:

“(...) it is a conceptual condition or presupposition of my intention to do my part of X that we indeed are about to participate in a joint action”
Tuomela and Miller (1988), p. 373.

Clearly this does not serve as a reductive definition, since it is circular. Remember that collective intention is the first step of our conceptual ladder. We can include neither a collective action X nor a collective agent G in the analysis of its definition without circularity. Tuomela acknowledges that, and so I think that we should construe his proposal not as a reductive definition but as a reflexive, “fix-point,” definition of a primitive concept of collective intention.

Bratman sees no vicious circularity in his own proposal. As he states in the quote above he makes his proposal as a reductive attempt. Then, he seems to belong to group A. But does he succeed in reducing collective intentions to singular ones? Let us examine his analysis.

Actually, it is not fair to represent Bratman's account by the graphic corresponding to group A. The intentions attributed to the agents participating in a collective action are, according to Bratman, not simply

I intend

but

(*) I intend that we do J,

where J is a collective action. On the face of it, this does not seem very different from a Searlean collective intention of the form 'We intend to J,' which is also an intention of an individual. According to Velleman's remarks, these different approaches would share the same strategy:

“to imagine two or more agents as individually holding different token intentions of the same type, by holding intentions formulated in the first-person plural. The idea is that you and I can partake in the same intention if there is something that each of us individually intends that “we” are going to do.” Velleman (1997), p. 33.

But there are, I think, important differences. First, Bratman's account

“exploit the fact that we speak not only of intentions *to*, but also of intentions *that*—for example, my intention that Scott clean up his room. Accordingly, we can speak of my intention that we J.” Bratman (1993), p. 102.

These ‘intentions *that*’ are not without problems. But let us assume that one can intend that we J.¹² He includes “we” and “J” (a collective action) in the definition of this intention. This can be circular. And Bratman acknowledges it:

“The content of this intention [that we dance a tango together] seems already to bring in the idea of shared intentional activity. But I wanted to explain shared intentionality by appeal to the idea of shared intention. So I worried that we were threatened by an unacceptable circularity.” Bratman (1997), pp. 51-52.

This is precisely one of my most important objections to Bratman's account: including a collective action J in the object of the collective intention he incurs in an unacceptable circularity. He proposes the following way out: According to him, in the case of singular action we avoid a similar circularity concerning intentionality by eschewing

“appeal to attitudes that include in their content the very idea of intentional action. Instead, we limit ourselves to contents that, at most, involve a concept of action that leaves it open whether the action is intentional.” Bratman (1997), p. 52.

Similarly in the case of his ‘intention that we J,’ he proposes that

“we try to understand J, as it appears in the content of my intention, in a way that leaves it open whether our J-ing is a case of shared intentionality.” Bratman (1997), p. 52.

Notwithstanding, we should notice that there are different types of circularity involved here. In the case of singular action, he is talking about the so-called causal self-reference of intentions. Is it my intention to raise my arm, an intention to raise it intentionally? Are we not including the very idea of intentional action in the content of intention, when our

¹² See Bratman (1997) and Postema (1995) for some of these problems and plausible responses. Anyway, ‘intentions that’ weaken Bratman's claim that in his account

“the idea is not to introduce some fundamentally new and distinctive attitude. The attitude we are appealing to is intention” Bratman (1997), p. 102.

This claim has to be weakened, especially when he contrasts it with Searle's collective intention. Collective intentions are also intentions and, after all, they do not seem to be more primitive (new and distinctive) than intentions *that*.

strategy is to define intentional actions via intentions? As he said in the quote above, in this case we can leave it open whether the action is intentional.¹³

However, in the case of collective intentions there are two sorts of circularity involved: one coming from the intentionality of J, and the other from the sharedness, collective nature of J. The first is harmless, but what about the second? If we leave it open whether J is a collective action, how does our ‘intention that we J’ characterize J as a collective action? One possible answer would be by the inclusion of ‘we’ in the content of my intention, but that would be again clearly circular. When we consider Bratman’s ‘intend that we J’ we should

“assume that we will have available appropriate conceptions of joint activity that are neutral with respect to shared intention; or anyway, my discussion is limited to such cases.” Bratman (1993), p. 101.

So, Bratman’s intentions do not make any appeal to any previous idea of collective action or agency. What are, then, the attitudes of the agents that make their action collective? This is his proposal of a definition of shared intention:

“We intend to J if and only if
1. (a) I intend that we J and (b) you intend that we J
2. I intend that we J in accordance with and because of 1a, 1b, and meshing subplans of 1a, 1b; you intend that we J in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b.
3. 1 and 2 are common knowledge between us.”
Bratman (1993), p. 106.

This “interlocking web of intentions” together with the common knowledge about it is what constitutes Bratman’s shared intention. Let us see whether it works. Remember that it has to distinguish between collective action proper and the sum (even if it is complex) of singular actions. Taking again our example, do the agents satisfy 1, 2, and 3? Action J will be ‘hit the policeman,’ with no collectivity built-in, and the same for ‘we’ in ‘intend that we J’. (1a) is then equivalent to the conjunction of

(1a’) I intend that I J (which, I guess, means just ‘I intend to J’)

and

¹³ Or even we can just talk about bodily movements and represent intentions in the following way:

I intend (raise my arm, as a result of this intention)

This is circular. But the circularity is harmless. There is no infinite regress. Harman (1976, 1986) and Searle (1983), argue for the causal self-referentiality of intentions. In contrast, Bratman suspects that intentions are causally self-referential only in the case of the intentions in “shared intentional activity” or collective action (Bratman 1992).

(1a'') I intend that you J.

So, if I intend to hit the policeman and I intend that you also hit him, I satisfy (1a), and similarly for you and (1b). It is not unnatural to think that, in the real case, the guys had just this type of intentions. They had each a singular intention to hit the policeman, and also an intention that the others do the same in order to get, for instance, the anonymity that provides acting within 'the mass.' And this can also explain that I intend that we J because (1a) and (1b); that is, if I intend that we J, then I intend that we J because I intend that we J (causal self-referentiality of intentions),¹⁴ and also I intend that we J because you intend that we J. In other words, my intention that we J is dependent on your intention that we J, because I would not intend that we hit the policeman if I knew (or believed) that you did not intend that we hit the policeman.

We can also think that their intentions are in accordance and because of their "meshing subplans." This condition does not mean, following Bratman, that their subplans should completely match nor that they should be fully determined, but what is required is that "each of us *intends* that we J by way of meshing subplans" that guarantee appropriate interpersonal coordination. All this can be common knowledge among us. So our action would satisfy all conditions in Bratman's definition. Yet our action would not be collective in the required sense. It would not count as an action of a collective, but as the aggregation (though coordinated) of our singular actions.

Hence, Bratman's definition in his intended interpretation—i.e., without any idea of collectivity incorporated in the action J or the pronoun 'we'—does not give us a correct analysis of collective intention. Bratman's analysis falls short unless it incorporates an irreducible collective sense of 'we' or 'J.' But thus it would be a circular analysis, failing then to reduce collective intentions to singular ones.

So far, we have one good reason to take collective intention as a primitive concept, in the sense suggested by Searle and also by Tuomela. But what do we mean by "primitive"?

In the same way that we say that personal intention is a primitive concept because it is not reducible to beliefs and desires, we say that collective intention is primitive because it is not reducible to singular intentions and mutual beliefs. We are saying that among the primitive notions we have for designating mental states we should add 'collective intention' to the list of belief, singular intention, desire, goal, etc.¹⁵ We mean that in our conceptual analysis

1. Collective intentions
- ↓
2. Collective action
- ↓
3. Collective agent, group, team, we...

¹⁴Bratman accepts that for this kind of intentions (see precedent note).

¹⁵ As we said before, we limit ourselves to the case of collective intentions. See note 9.

we cannot take a previous step and analyze collective intentions in terms of other kind of singular mental states. It does not mean that we are accepting supra-individual minds and thus rejecting methodological individualism. Remember that in spite of their name collective intentions are *individual mental states*. So, how can we explain the existence of collective agents? We will consider it in the next section.

5. Constructing us

The clearest account of the existence of collectives that is in accordance both with methodological individualism and the primitive nature of collective intentions is, in my opinion, Searle's (1995, 1997):

“On my view, the existence of collective intentionality as a psychological primitive in the individual heads of individual agents does not commit one to a primitive ontology of actual human collectives. On the contrary, the basic ontology is that of individual human organisms and their mental states. The collective arises from the fact that collective intentionality is in the individual heads of individual organisms. The actual social collective consists entirely of individual agents with collective intentionality in their heads, nothing more. Ontologically speaking, collective intentionality gives rise to the collective, and not the other way around.” Searle (1997), p. 449.

So we begin with collective intentionality, with collective intentions in our case, intentions that are characterized by an irreducible *we-ness*. It is having these intentions what makes possible the performance of a collective action by a set of individual agents, which constitute now a collective agent. So, this need not be considered as an ontological primitive:

“My conclusion is that social collectives can be constituted by the fact that individual agents think of themselves as part of a collective without thereby supposing that the collective is an ontological primitive. The collective's existence consists entirely in the fact that there is a number of individual agents who think of themselves as part of the collective.” Searle (1997), p. 450.

On the one hand, we have an explanation of the existence of collectives, which does not consider them ontologically primitive. On the other hand, we need a primitive notion of collective intention. This sort of intention is still as individual as any other mental state, and thus it is consistent with the hypothesis of the brain in the vat:

“I could have all the collective intentionality that I could want in my head and still be radically mistaken. The fact that I have a *we-intention* does not

by itself imply that other people share my we-intention, or even that there is a “we” that my intention refers to. I take myself to be engaging in collective behavior with other people, but whether or not I am in fact succeeding in engaging in collective behavior with other people is not a matter of the contents in my head. The existence of collective intentionality does not imply the existence of human collectives actually satisfying the content of that intentionality. But once you have collective intentionality, if it is *in fact* shared by other people, the result is more than yourself and other people: collectively you now form a group.” Searle (1997), pp. 449-450.

A collective intention, then, is not an intention of a group, but an intention that individuals have if they form a group. This is another difference between our Searlean notion of collective intention or Tuomelian notion of we-intention and Bratman’s shared intention:

“Shared intention is an intention of the group. You and I together may have a shared intention to dance a tango. *I* cannot by myself have a shared intention to dance a tango, though I can intend to dance one with you” Bratman (1997), p. 50.

But what does it mean “intention of a group”? Might he be talking about supra-individual mental states, about group minds? I do not think so. He adds just immediately:

“On my proposal I can also have an intention that *we* dance the tango; and it is such an intention on my part, together with an analogous intention on your part, which is central to *our* shared intention to dance the tango.” Bratman (1997), p. 50.

The difference could be seen just as a terminological one. Searle’s “collective intention” and Tuomela’s “we-intention” stand for the intentions in each individual of the group, whereas “shared intention” describes a state of the group of agents when each of them has an intention “that we J.” Bratman’s term is more appropriate since for something to be *shared* or *collective* it seems that more than one agent is required. However, “shared intention” is still misleading because it is not really an intention but the conjunction and interrelation of the intentions of the agents in the group. Actually, the different terms point to a difference on perspective. Bratman is describing the group and the agents’ mental states from an objective, God-eye like perspective. Searle, on the other hand, adopts a subjective perspective. The two perspectives are not incompatible, but it is important to make clear which one we are adopting for not viewing nonexistent differences, and being able to detect the real ones.

By now, we have argued for a primitive notion of collective intention, i.e. non-reducible to any configuration of singular intentions and mutual beliefs, which stands for an *individual* mental state, and make justice to the idea of collective action as essentially different for singular action without violating the thesis of methodological individualism.

The crucial question now seems to characterize collective intentions by their properties and by their relations with other mental states.¹⁶ But before doing that, we should face an important objection to these analyses of collective intention, which alludes to an alleged constraint on the possible objects of intention.

6. The ownership and control of *our* actions

Traditionally, philosophers of action have considered that the object of one's intentions could only be one's own actions; one could only intend actions that are within one's power to do or bring about:

“The proper objects of intending... seem limited to my actions (not the sun's) and to things I can do.” (Baier (1970), p. 649.

That may seem reasonable, considering intentions as individual mental states that cause that one's bodily movements constitute an action. The cause of my action is my intention; the cause of your action is your intention. When I have a collective intention, that is, I intend an action of *ours*, am I violating this basic condition on intentions? If so, there could not be, by principle, any intention formulated in the first person plural, any collective intention.¹⁷

In fact, I think this can be a more serious objection when collective intentions are analyzed à la Bratman. As we have seen, his analysis involves intentions of the type “I intend that you J,” and, therefore, taking as the object of one's intention the action of yours, would violate the ownership condition of the objects of intention. Indeed, within the mental cause theory of action would be bizarre to see my intention as the cause of your bodily movements. However, this kind of intentions is not necessary in the analysis of collective intentions. I suggest that one's collective intention has the same action-motivating power as one's singular intention. For instance, when I have the collective intention that we (you and me) push the car, I do not need any other intention, in particular any other singular intention to start my bodily movements. For us to push the car it is necessary that *I* push the car; in that sense my collective intention is enough for explaining my behavior. Do I need to intend that you also push the car and violate the above condition? No, what happens when I have a collective intention that we push the car is that I also believe that you have the corresponding collective intention. Or, further, I probably believe that is mutually believed between you and me that each has a collective intention to push the car. If I am right in my (mutual) beliefs, you will have

¹⁶ Miller and Tuomela (1999) propose a classification of goals with respect to their carriers and contents. I don't think this makes much sense when talking about full-blown collective intentions.

¹⁷ This is the main subject addressed by Bratman (1997) and Postema (1995). See their arguments for allowing the possibility of collective intentions. Anyhow, Bratman's solution involves stipulating a distinction between *intentions to* and *intentions that*. I agree with Postema that “the plural perspective is not necessarily committed to impossible intentions.” Postema (1995), p. 60. My arguments, however, do not rely on his.

your collective intention to push the car, and that will be the cause of your action: your intention, not mine.

The own action condition then, seems to pose no problem on collective intention, because when I have a collective intention I am not intending that *you* intend, I only believe (mutually with respect to you) that you already have a collective intention of the same type. Moreover, when we consider conditions in a Bratmanian sense as future-directed intentions, collective intentions have the same characteristics of singular intentions:

“Our J-ing may serve as an end for my planning and my actions. I may constrain my plans so as not to settle on options incompatible, given my beliefs, with our J-ing. I may see myself as faced with a problem about how, given the end of our J-ing, this is to be achieved. Our J-ing may, that is, serve as an end in my plans, constrain those plans, and pose problems of means and preliminary steps for those problems.” Bratman (1997), p. 53.

So, in these points collective intentions play the same characteristic role of singular intentions. And, as Bratman concludes, we are still without an argument showing that there is an intrinsic incoherence in this kind of intentions.

What are then the characteristic properties of collective intentions? We will devote the last section to the discussion of some of the most important ones.

7. Some properties of collective intention

For studying the properties of collective intention it can be convenient to use a formal notation. We will use the formula

$\text{COLINT}_{xy} \alpha$

which can be read in a Bratmanian vein as

“I (x) intend that we (y and me, collectively) do α ”

or in a Searlean way as

“We (y and me, collectively) intend to do α .”

The important point is that they are intentions of the individual x , which subjectively characterizes α as a collective action, and x and y as a collective agent, as *we*. Now, from an objective point of view, for α to be a truly collective action, it must be true that α has been performed under the following conjunction:

$$\text{COLINT}_{xy} \alpha \wedge \text{COLINT}_{yx} \alpha.$$

When Bratman talks about the intention of a group, this seems to be what he is talking about. But now it is obvious that this is not an intention, not even a shared intention, but the conjunction of the collective intentions of the members of the group. Since what we want to clarify are the properties of this type of mental states and its relationship with other mental states of the agents, we adopt the subjective perspective.

In the last section, when talking about the beliefs associated to one's collective intention, we were pointing to one particular feature of collective intentions: their overt nature. Traditionally, singular intentions can be grouped into three classes in terms of the relationship between their satisfaction and their recognition by other agents: those which are neutral; those which are covert, i.e., intentions that are not to be recognized by other agents to be satisfied; and those which are overt, namely, whose recognition is necessary for their satisfaction. Communicative intentions are a typical case of the last class. Well, collective intentions are also overt.¹⁸ Tuomela and Miller and also Bratman clearly are assuming that with the last clause in their respective analyses.

As we said before, when I have a collective intention to push the car I also believe that you have the corresponding collective intention. In our formal notation:

$$\text{COLINT}_{xy} \alpha \rightarrow \text{BEL}_x \text{COLINT}_{yx} \alpha.$$

But probably is not only a simple belief on my part concerning your collective intention. Due to the overt nature of collective intentions, I have to believe that you believe that I believe you have that corresponding collective intention. In other words I have the mutual belief (with respect to you) that you have that collective intention. Namely,

$$\text{COLINT}_{xy} \alpha \rightarrow \text{MB}_{xy} \text{COLINT}_{yx} \alpha.^{19}$$

This captures the overt nature (from my subjective point of view) of your collective intention. But, surely my collective intention is also overt. So we would have:

$$\text{COLINT}_{xy} \alpha \rightarrow \text{MB}_{xy} (\text{COLINT}_{xy} \alpha \wedge \text{COLINT}_{yx} \alpha).$$

This formula is intended to represent the fact that when I have a collective intention to do α with you, I assume that there is a reciprocal intention on your part, and that these intentions are overt between us. Of course, collective intentions are overt for the agents in

¹⁸ The difference between communicative intentions and collective intentions concerning their overt nature is clear: for communicative intentions their recognition is necessary and sufficient; for collective intentions it is necessary but not sufficient.

¹⁹ Where $\text{MB}_{xy} p \equiv \text{BEL}_x (p \wedge \text{MB}_{yx} p)$. This is the fixpoint version of one-sided mutual belief (Colombetti 1993). Again, we are interested in operators that represent mental states of the agents. $\text{MB}_{xy} p$ is a belief of x , and it is compatible with the nonexistence of the corresponding belief of y , that is, with $\neg \text{BEL}_y p$ and even with $\neg \text{MB}_{yx} p$.

the collective action, but with respect to other agents outside the group it could be either overt or covert, or neutral.²⁰

It is important to note that my having a collective intention implies neither that you also have it nor that you believe I have it. In other words the formula above is consistent both with

$\neg\text{COLINT}_{yx} \alpha$

and

$\neg\text{BEL}_y (\text{COLINT}_{xy} \alpha)$.

So our concept of collective intention meets the condition of the brain in a vat.

We could go further on the description and analysis of the properties of collective intention with all the basic mental states of the agents involved in a collective action, and we could insert that analysis taking into account that intentions in general give rise to plans.²¹ That, no doubt, would lead us to a logic of collective intention within a KD45 logic for beliefs, $\text{KD4}_{bg}\text{5}_{bg}$ for goals and $\text{KD4}_{bi}\text{5}_{bi}$ for intentions in the case of singular agency plus, specifically, a $\text{KDD}^*4_{bci}\text{5}_{bci}4_{mbci}4_{mbci}$ for collective intentions.²²

²⁰ Related to the overt nature of collective intentions is what Tuomela calls the Collectivity Condition of satisfaction for collective intentions:

“If my intention that we perform X together is satisfied, then, on quasi-conceptual grounds, also your intention that we perform X together is satisfied, and vice versa ... and each of us believes so.” Tuomela (1998), p. 137.

²¹ Future directed intentions function, according to Bratman, as input to practical reasoning. The adoption of an intention poses a planning problem for the agent. She has to adopt a plan to satisfy her prior intention. In these cases, the agent ‘s collective intention will imply an intention of the following type:

$\text{COLINT}_{xy} \alpha \rightarrow \text{INT}_x \beta$, such that $\beta \in \alpha$ and the relation \in is defined in this way: $\beta \in \alpha$ if and only if β enables α and $\beta \cup \{\delta_i\}$ generates α . (The set $\{\delta_i\}$ corresponds to the collection of actions that constitutes a plan to do α .)

In short, according to our analysis, x ’s having the collective intention to do α involves, the intention to do her part of the collective action α . This formulation captures, I think, both Tuomela and Miller’s first clause of we-intentions and Bratman’s reference to each participant’s subplans of the joint action. I do not think that the agent has always to adopt a later intention of this kind. The collective intention to push the car can be the only intention who causes her pushing as part of our pushing, but there can be other actions more complicated and expanded over time that asks for this division of labor. Miller and Tuomela’s (1999) classification of goals can be relevant but I do not pursue the matter here.

²² See Korta and Larrazabal (2001). The axioms, in modal logic, are:

K. $\text{COLINT}_{xy} (\alpha \rightarrow \beta) \rightarrow (\text{COLINT}_{xy} \alpha \rightarrow \text{COLINT}_{xy} \beta)$

D. $\text{COLINT}_{xy} \alpha \rightarrow \neg\text{COLINT}_{xy} \neg \alpha$

It seems evident that the clarification of so important a concept as that of collective intention can only be achieved by a deep interaction among a good conceptualization, a clear logical proposal of a calculus, and a rigorous semantics. This is the way to follow, in my view, and probably we are just at the beginning, in need of testing this framework in such different domains of application as systems of collective agency in Distributed Artificial Intelligence, pragmatic analysis of discourse, and collective agency in social psychology. That is the only way to avoid what seems to be the current greatest danger of “collectivizing” all psychological notions, mixing confusedly psychology with sociology and, by the way, not distinguishing what is a mental state and what is not.

8. References

- Baier, A. (1970), Act and Intent. *The Journal of Philosophy* 67, 648-658.
- Balzer, W. and R. Tuomela (1997), A fixed point approach to collective attitudes. In G. Holmström-Hintikka and R. Tuomela (eds.). *Contemporary Action Theory. Vol. 2: Social Action*. Dordrecht: Kluwer, pp. 115-142.
- Bratman, M. E. (1992), Shared Cooperative Activity, *The Philosophical Review* 101(2), 327-341.
- Bratman, M. E. (1993), Shared Intention. *Ethics* 104, 97-113
- Bratman, M.E. (1997), I intend that *we J*. In G. Holmström-Hintikka and R. Tuomela (eds.). *Contemporary Action Theory. Vol. 2: Social Action*. Dordrecht: Kluwer, pp. 49-63.
- Colombetti, M. (1993), Formal semantics for mutual belief, *Artificial Intelligence* 62, 341-353.
- Gilbert, M. (1987), Modelling Collective Belief. *Synthese* 73, 185-204.
- Gilbert, M. (1989), *On Social Facts*. London: Routledge.
- Gilbert, M. (1997), What is it for us to intend? In G. Holmström-Hintikka and R. Tuomela (eds.). *Contemporary Action Theory. Vol. 2: Social Action*. Dordrecht: Kluwer, pp. 65-85
- Harman, G. (1976), Practical reasoning. *Review of metaphysics* 29, 431-63.
- Harman, G. (1986), *Change in view*. Cambridge, MA: MIT Press.
- Hobbs, J. R. (1990), Artificial Intelligence and Collective Intentionality: Comments on Searle and on Grosz and Sidner. In P. R. Cohen, J. Morgan and M. E. Pollack, *Intentions in Communication*. Cambridge, MA: The MIT Press, 1990, pp. 445-459.

D*. COLINT_{xy} α → ¬INT_x ¬ α

4_{bci}. COLINT_{xy} α → BEL_x COLINT_{xy} α

5_{bci}. ¬COLINT_{xy} α → BEL_x ¬COLINT_{xy} α

4_{mbsci}. COLINT_{xy} α → MB_{xy} COLINT_{xy} α

4_{mbsci}'. COLINT_{xy} α → MB_{xy} COLINT_{yx} α

- Korta, K. and J.M. Larrazabal (2001), On the logic of collective intention. *The Bulletin of Symbolic Logic* 7(1), March 2001, 126-127.
- Miller, K. and R. Tuomela (1999), What are collective goals? Draft.
- Popper, K. (1969), *The Open Society and its Enemies*. London: Routledge.
- Postema, G. (1995), Morality in the First Person Plural. *Law and Philosophy* 14, 35-64.
- Searle, J. (1983), *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Searle, J. (1990), Collective intentions and actions. In P. R. Cohen, J. Morgan and M. E. Pollack, *Intentions in Communication*. Cambridge, MA: The MIT Press, 1990, pp. 401-415.
- Searle, J. (1995), *The Construction of Social Reality*. London: Penguin.
- Searle, J. (1997), Précis of *The Construction of Social Reality*. *Philosophy and Phenomenological Research* 57, 427-458.
- Stoutland, F. (1997), Why are Philosophers of Action so Anti-Social? In L. Alanen, S. Heinämä and T. Wallgren (eds.), *Commonality and Particularity in Ethics*, London: Macmillan, pp. 45-74.
- Tuomela, R. (1995), *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford: Stanford University Press.
- Tuomela, R. (1998), Collective Goals and Cooperation. In Arrazola, X. et al. (1998), *Discourse, Interaction and Communication*. Dordrecht: Kluwer.
- Tuomela, R. and M. Bonnevier-Tuomela (1997), From Social Imitation to Teamwork. In G. Holmström-Hintikka and R. Tuomela (eds.). *Contemporary Action Theory. Vol. 2: Social Action*. Dordrecht: pp. Kluwer, 1-47.
- Tuomela, R. and K. Miller (1988), We-intentions, *Philosophical Studies* 53, 367-389.
- Velleman, J. D. (1997), How to share an intention. *Philosophy and phenomenological research* 57, 29-50.