

Chapter 2

Overcoming the ‘Cartesian Brainwash’

Beyond Intentional Individualism

Among the many reasons why John Searle is important in the debate revolving around the structure of collective intentionality is the fact that even though the history of the analysis of collective intentionality has roots that go further back,¹ it was him who coined the term (Searle 1990). The following chapter sheds some critical light on a feature which Searle’s account shares with most of the received accounts of collective intentionality. It is argued that fear of the group mind has played a fateful role in the early stages of the current debate by driving most philosophers of collective intentionality into accepting one or another version of intentional individualism.

§6 Collective Intentionality Without Collectivity?

On his way towards a “*general theory*” (Searle 1998a: 161), John R. Searle has recently started to venture into what he likes to see as a new field: ‘Philosophy of Society’. In some of the papers surrounding his *Construction of Social Reality*, Searle envisages this discipline to be centered on how the individual and society relate to each other (1997b: 103, 1998b: 143). Of course, this is hardly a new question. It has been the topic of many a philosophical debate and controversy at least since it became common practice to refer to single human beings as individuals. In the second half of the last century, individualism has become the dominant view of the basic structure of this relation. In much of social science, it has come to be widely held that explanations of social phenomena have ultimately to be given in terms of individual actions (cf. Popper 1962: 98). Let us label this the orthodox view. In its reductivistic form, the social is *nothing but* an aggregate of individuals who decide over the alternatives they believe to be available to them in the light of whatever preferences they have; in the last resort, it is each individual deciding for him- or herself. Thus the social is *secondary* as compared to the intentionality of the single

¹ The direct roots of the concept are in Robin George Collingwood’s *New Leviathan* (1942), where Wilfrid Sellars picked up the term, which was then analyzed by Raimo Tuomela (1984; see also Tuomela/Miller [1988]) who initiated the current debate.

individuals. It appears that we do not have to presuppose collectivity concepts such as 'group' or 'community' in order to analyze what it means for an individual to optimize his or her expected utility. Collectivity concepts enter orthodox explanations of social phenomena only insofar as they are either the direct *object of individual intentions*² or among the *unintended consequences of individual actions*.³ In both cases, individual intentions and actions – and *not* collectives – are what social science is about. For in this view, the social does not reach down to the form and structure of intentionality and action itself.

In spite of the near total "Triumph of the Individual" in social science, some opponents and contesters survive. In philosophy, it is widely held that interpretation (and thus interaction) is a precondition for there to be mental states with intentional content. Social externalism even seems to have become the mainstream position. In social and sociological theory, too, some 'heterodox' strands persist in opposing the allegedly 'atomistic' picture of human agency and intentionality. Heterodox philosophers of society emphasize that in many respects, individual intentionality is more deeply imbued with sociality than the 'orthodox' view has it.⁴

At first glance, it might appear that the heterodox view receives further support from one of the most exciting recent developments in analytic philosophy of intentionality and action. The works of, among others, Raimo Tuomela, Margaret Gilbert, and Michael Bratman, together with John Searle's own contributions, have substantially broadened our understanding of intentionality and action. After the traditional concentration on the individual intentionality of single agents, the focus of attention has now shifted to an analysis of what it means to intend and act *together*, a phenomenon which, by and large, had received only marginal notice in the earlier philosophy of intentionality and action.⁵ By now, it is well-established and widely recognized as a fact that intentionality is not exclusively a matter of the personal beliefs, desires and expectations of individuals. What makes our intentionality and our actions *social* is not just that from time to time, we make each other the object of our individual intentions or expectations. Rather, intentionality is *in itself* something human beings can *share*.

This, it seems, adds a new facet to the question not only concerning the relation between the individual and the collective, but more generally concerning the relation between the mental and the social. In his *Common Mind* ([1992] 1996), Philip Pettit has distinguished two dimensions of that relation: the "vertical" issue concerning the question of whether or not collective forces trump intentional regularities, and the "horizontal" issue of whether or not – or how far – the capacities

² For a classical expression of this view see Weber ([1921] 1980: 7). For Weber, collectivities can be 'real' only in the sense that they are *believed to be real* by the individuals.

³ For this view see Elster (1989).

⁴ The most commonly known 'heterodox' strand in social ontology is Communitarianism; cf. Sandel (1982).

⁵ Early exceptions to the rule can be found in Phenomenological Philosophy (see, e.g., Walther (1923)); for a more "holistic" view see Stein (1922: 116–267); Heidegger ([1928/29] 1996: 83ff.); Sartre ([1943] 1991: 464ff.).

that mark us out as human beings essentially depend, for their existence, on social relation. Pettit proposes to label the first issue the controversy between individualism and collectivism, and to use the Greek counterparts of these terms – atomism and holism – for the contending positions in the second issue. Famously, Pettit argues for individualism (and against collectivism) in the vertical dimension, and for holism (and against atomism) in the horizontal dimension. But he does not, in this book, address the particular way from which Tuomela (1984, 1995), Searle (1995), Gilbert (1989) and Bratman (1999) had started to approach the question of the social just a few years before Pettit's *Common Mind* first appeared in print. Comparing the way in which Pettit treats the question of the “commonality” of the mental with the way in which these other philosophers approach the topic makes the novelty of the analysis of collective intentionality all the more apparent. Pettit addresses the traditional issues that have been on the agenda of research in social ontology at least since social externalism first appeared on the scene; with Searle et al., a new perspective has opened up – a perspective that largely seems to be independent of the earlier debates. In order to say that intentionality can be genuinely *shared* one need not assume that metaphysical socialism about intentionality is true, i.e. that *any kind of intentionality* is a social fact, let alone that collective forces compromise the intentional psychology of individuals as some extreme collectivists had it.⁶ Thus collective intentionality does indeed open up a new perspective on the relation between the individual and the collective levels.

Upon a closer look, however, it appears that heterodox philosophers of society should not put their hopes for support from the theory of collective intentionality too high. The main protagonists of this movement do not seem to think that their novel approach to the structure of intentionality and action should open a new perspective on the basic structure of the relation between the social and the individual. By and large, the orthodox account is left intact. Raimo Tuomela virtually treats groups as ontological non-entities because in his view, “groupness” is, as he puts it, “in the last analysis attributed to individuals” (Tuomela 1995: 199). It seems that in his account, the basic structure of we-intentionality does not *per se* presuppose collective entities such as ‘groups’ or ‘communities’. For an individual to we-intend it is, following Tuomela, not necessary that other agents actually exist, much less that there is an actual we-group.⁷ Margaret Gilbert, meanwhile, has repeatedly claimed to go “beyond individualism” (Gilbert 2000: 3). Yet in her book *On Social Facts*, she explicitly bases her analysis on a concept of the individual that “does not require for its analysis a concept of a collectivity” (Gilbert 1989: 435ff.). The conceptual basis of her account of “joint commitment” consists of nothing but conditional personal commitments (Gilbert 2002). Michael Bratman, in turn, calls his own theory “reductive in spirit” because he takes shared intentionality to be analyzable “in terms of attitudes and actions of the individuals involved” (Bratman 1999: 108). Last but

⁶ In the meantime, Pettit has taken his stance on that new topic (Pettit 2005). It seems, however, that Pettit does not recognize that the construction of collective agents presuppose collective intentional activity from the side of the participants.

⁷ Tuomela (1991: 254). See also Hindriks (2002).

not least, Searle himself not only hastens to declare that his account of collective intentionality is fully consistent with methodological individualism (Searle 1990: 406). He also stresses the ontological primacy of what goes on in the individual mind over the existence of the group by pointing out that, "ontologically speaking, collective intentionality gives rise to the collective, and not the other way around" (Searle 1997a: 449).

Throughout this debate, actual collectivity is, it seems, held to be methodologically and ontologically secondary to (and derivative from) the mental activity of the single individuals involved in collective intending. The actual existence of a we-group is seen as a more or less contingent by-product of the intentionality of individuals. From a heterodox point of view, this debate gives the impression of an attempt to account for the structure of collective intentionality without letting any genuine collectivity enter the scene. "We-ness" is the topic, yet at the same time it is stressed that it is a feature of individuals – and not of an actual 'we'. Annette Baier expresses the dissatisfaction heterodox philosophers of society might feel rather drastically when she takes the current debate on collective intentionality to prove that Descartes has thoroughly brainwashed us (Baier 1997b: 18).

§7 The Specter of the Group Mind

Before taking a closer look at this sweeping diagnosis, I would like to highlight a rather somber figure that is haunting this debate, and that seems to have played a crucial role in the formation of its individualistic setting. It is the specter of the *collective subject*, or *group mind*. Its importance in this debate seems to stem from a rather innocent-looking assumption. Where there is intentionality, it is said, there has to be somebody who 'has' it – the good old subject. Now if it is claimed that there is such a thing as *collective* intentionality, and that collective intentionality is to be distinguished from *individual* intentionality, the conclusion seems to force itself on us that it has to be, not *single individuals*, but *collectives* themselves that 'have it'. And for collectives to have intentions, some sort of a 'collective mind', some 'group mind', seems to be required, something hovering over and above the minds of the individuals involved. To the untrained eye, at least, this apparent implication of the very concept of collective intentionality does not look very appealing. Thus, among the protagonists of collective intentionality, it was originally widely agreed upon that there is no mind over and above the minds of individuals (ironically, it was Pettit who eventually showed that one need not be all that worried about ascribing a "mind of their own" to groups [2003]). While the question of whether or not (and, if so, in what sense) collectives can be agents in their own right remained to some degree controversial even in the early analyses, it seemed almost universally agreed upon that it is unacceptable to treat collectives as 'subjects' of intentions and actions in the *ordinary* sense in which individuals are the bearers of

their intentionality.⁸ Even where the notion of the collective subject was stripped of its mentalistic content, it still did not quite appeal to most philosophers of mind and action, because it seemed to be associated with collectivistic⁹ (or even totalitarian¹⁰) notions of the social. If it is to the *collective* rather than the individuals that intentionality is ascribed in order to make sense of an observed behavior, the participating individuals seem to be no more than organs, i.e. mere instruments, and this seems to contradict our idea of individual intentional autonomy.

Thus it seems quite understandable that the above-mentioned philosophers of collective intentionality set themselves the task of showing that collective intentionality is possible without there being a group mind (let alone some collective consciousness) involved. The specter of the group mind (or collective subject) had to be exorcised, and one can identify two different ways in which this was done. The softer way – it might look more like psychotherapy than like hard-core exorcism – was chosen by Margaret Gilbert, Raimo Tuomela, and, perhaps, Robert Sugden. In these analyses, some sort of collective subject is admitted to the theory, but it is domesticated so as to be consistent with an otherwise thoroughly individualistic conceptual framework. Here, either some rather strong sense of membership to a collective (Sugden¹¹) or some softened and modernized version of the collective subject itself (Gilbert,¹² Tuomela¹³) is made part of the theory. At the same time, however, the collective subject is solidly founded in the intentional autonomy of individuals by reducing the collective subject either to sets of individual intentions¹⁴ or to the reflective self-understanding or self-categorization of the single participating individuals qua members of the team.¹⁵ The tougher way of dealing with the specter of the group mind was simply to treat it as an abominable collectivist idea that has to be banished from the theory of collective intentionality straight away. On this tough line, the group mind is exorcised either by stating that all intentionality involved in collective intending is exclusively the intentionality in the minds of the participating individuals, or by making the somewhat different point that the intentionality individuals “have” when participating in collective intentionality is basically a form of their personal intentionality. These are the strategies that were chosen by Searle and Bratman, respectively.

⁸ Edmund Husserl’s theory of “higher order-persons” gives an illustrative example of the difficulties that any attempt to apply the model of the individual subject to collectives will face (cf. Schmid 2000: 17–27).

⁹ Cf. Emile Durkheim’s concept of the “collective consciousness” ([1898] 1994).

¹⁰ Cf. Hartshorne (1942).

¹¹ Sugden favors a concept of membership “in something like the old sense in which arms and legs are members of the body” (1993: 86). This reminds of the Aristotelian view of the relation between society and the single human beings (see Aristoteles: *Politics* 1253a), a view that – at least at first glance – appears to be inconsistent with our modern view of the single human beings as *individuals*.

¹² Cf. Gilbert’s concept of the “plural subject” (1989).

¹³ Cf. Tuomela 1995: 231).

¹⁴ See Tuomela (1995).

¹⁵ See Gilbert’s “Simmelian” account in Gilbert (1989: Chap. 4). In Sugden’s view, “a team exists to the extent that its members take themselves to be members of it”; Sugden (2000: 192).

Before turning to this, let me state a general observation. Looking from afar at how the group mind was dealt with in the early stages of the debate on collective intentionality, it might appear that the theory of collective intentionality was caught in a dilemma, or rather, stuck in some kind of double-bind, as it were. On the one hand, the aim was clearly to break with individualism in the sense of the orthodox limitation to purely individual intentionality, which is recognized as being overly restrictive and unfit for our understanding of the social world. On the other hand, however, individualism (in the broad sense of an emphasis on the role of the individual) seemed to be the only effective defense against the specter of the group mind. Thus, in a sense, the theory of collective intentionality had to reject and to endorse individualism at the same time. How was this apparent tension dealt with? In exploring this question further, and in showing how detrimental this constellation was for the further development of the analysis of collective intentionality, I shall concentrate on the hard-line accounts against the group mind, i.e. Searle's and Bratman's.

Following up on Baier's statement concerning the 'Cartesian brainwash', a short remark on Descartes is in order. What is 'Cartesian' about how such authors as Searle and Bratman put their respective analyses of collective intentionality? Let us start with the most obvious sense in which there is something non-social about Descartes' venture. In his *Meditationes*, Descartes makes clear that his aim is to contemplate his own mind in "lonely withdrawal"¹⁶ from society. Thus it is hardly surprising that he comes up with a rather under-socialized account of the mind. There are two ways in which his view of the mind is individualistic. Firstly, the mental comes exclusively in the form *ego cogito* – and not, as Charles Horton Cooley would have already liked to have it, in the form *nos cogitamus*.¹⁷ This is to say that Descartes' account is individualistic in that it restricts intentionality to the form "I intend", "I think". It simply does not seem to have crossed Descartes' mind that there could be intentionality in the first person *plural* form, too. I shall refer to this first version of individualism with the term *formal individualism*, for what is at stake here is the *form* of intentionality.¹⁸

In a second and quite different sense of the term, Descartes' account is individualistic in that he portrays the individual mind as a *solitary place of representations*. Whatever the contemplating self finds in its mind is, following the view that was first articulated by Descartes, *structurally independent* of any relation to anything outside that individual mind. There is no telling whether a belief does or does not represent a real state of affairs just by reflecting on that belief *qua* mental state. Even the existence of some *genius malignus* who has the power of making me be mistaken in my beliefs could not thereby bring about the slightest structural change in my intentionality. "Being in a state with specific cognitive content does not essentially involve standing in any real relation to anything external" (Segal 2000: 11). In

¹⁶ Cf. Descartes, René: *Meditationes de prima philosophia*, First Meditation, §3.

¹⁷ Cf. Cooley ([1902/05] 1956: 6).

¹⁸ There are other terms in use for this kind of individualism. Kay Mathiesen (2002) proposes the term "phenomenological individualism" as opposed to ontological individualism.

the current debate, this view usually goes under the label ‘internalism’, but since internalism is usually taken to include a non-Cartesian account of the relation between the features of our physical brain and our mind, I shall use the term *subjective individualism* instead.¹⁹ This is the second sense in which Descartes’ view of the mind is individualistic. As opposed to formal individualism, subjective individualism does not limit intentionality to the singular *form*, but restricts the class of possible *subjects*, or ‘bearers’, or ‘owners’ of intentions to single individuals.

What is the role of this distinction between two versions of ‘Cartesian’ individualism in the current debate? As mentioned above, Bratman and Searle both reject individualism in breaking away from the orthodox standard model of intentionality and, at the same time, resort to individualism when they see themselves confronted with what they perceive to be the ugly face of the group mind. In this apparently paradoxical venture, the distinction between the two versions of individualism comes in handy: both Bratman and Searle choose to depart from *one version of individualism* in setting apart their respective concepts of collective intentionality from the standard model, and to resort to the *other version of individualism* in order to banish the group mind. Interestingly, however, they do not seem to agree on which version of individualism to throw out, and which one to keep! Bratman’s conception of shared intentionality seems to go beyond subjective individualism in some respects and to hold on to formal individualism, whereas Searle makes the opposite move. This results in a rather peculiar constellation: who is right? Or should it turn out that both are equally right (when they reject one form of individualism) and wrong (when they endorse the other form of individualism)?

Before turning to this question, let me establish the facts about Bratman’s and Searle’s respective forms of individualism. Bratman argues that what he calls “shared intention” or “group intention” is not anything single individuals can ‘have’ for themselves, but rather an “interrelation” (1999: 114) or an “interlocking web” (1999: 9) of what goes on in the minds of many individuals. Thus it seems clear that, in this account, “shared intentionality” cannot be structurally independent of external relations. We have to stand in actual relations for our intentionality to be shared. What makes our intentionality *shared* goes beyond the minds of single individuals. Thus Bratman seems to reject subjective individualism in the sense defined above. At the same time, however, Bratman deems it necessary to endorse formal individualism in order to not get stuck with the group mind. He hastens to declare that the relations presupposed in shared intentions are not tantamount to some “fusion” of individual agents to a “superagent” (1999: 111, 122ff.). In this respect, Bratman stresses that his account is thoroughly “reductive in spirit” (1999: 108). He rejects the idea that individuals literally *share* what they have in mind when intending together by emphasizing that the element of “we-ness” involved in what individuals intend when engaging in shared intentionality is reducible to a special

¹⁹ Another term that is in use for this view is “methodological solipsism” (see Searle 1990). In the given context, I find this term misleading, for the question at stake here is clearly not simply a question concerning methodology, but an *ontological* question concerning the subject or bearer of intentionality.

form of I-intentionality: intentions of the form "I intend that we J" on the part of the individuals involved, together with mutual knowledge of this intentionality and some matching relation between what the individuals intend, make up shared intentionality.²⁰

Significantly, it was not the interrelationistic move beyond individualism that has been most criticized in the debate of Bratman's account, but rather his reduction of we-intentions to sets of I-intentions (i.e. the move Bratman makes to avert the group mind). The upshot of a long discussion²¹ is the following. It seems that I have to take myself to be in a rather *influential position* within the we-group in order to form intentions of the form "I intend that we J" do.²² The reason is that intentions of this form extend into other people's intentional domains. Whoever has such intentionality in a sense intends other people's behavior. Thus he or she has to take others to be *responsive* to his or her own intentionality in a suitable way: he or she has to take his or her intentionality to be of some *influence* on other people's behavior (cf. remember the Principle of Intentional Self-Confidence as described above in chapter 1). Is this compatible with Bratman's account? My own impression (which I cannot argue for at length here) is that this "influence-condition" (1999: 116) shows that Bratman's account presupposes the element of sharedness it aims to explain.

Consider the following example. If we jointly intend to meet for lunch today, it does not seem necessary – indeed it is redundant – for me to form an intention of the form "I intend that we meet for lunch today" (rather, I will typically form some we-derivative [Sellars 1980: 99] or participatory [Kutz 2000a] intention of the kind "I intend to call you before noon to arrange a meeting place"). If and only if I take myself to be in a position to have a say in that matter, I might form an *additional* intention that *specifies* the content of our we-intention, and this additional intention might be of the form "I intend that we J_x" (e.g., "I intend that we have lunch together at the Japanese restaurant"). But intentions of this sort *presuppose* shared intentions instead of being their building blocks. It is only because *we intend J* that I can have intentions of the form "I intend that we J_x". Thus it seems that Bratman's "reductive" account of shared intentionality "in terms of attitudes and actions of the individuals involved" (1999: 108) simply fails to give an account of the crucial

²⁰ Cf. Bratman's conceptual analysis in Bratman (1999: 105).

²¹ See Baier (1997a/b); Stoutland (1997, 2002); Velleman (1997); Bratman (1999: 149–156); Kutz (2000a/b). Concerning the question of whether or not intention should be put in propositional or action-referential terms cf. the remarks above in Chapter 1.

²² At first glance, it might appear that intentions of the form "I intend that we J" are simply impossible. It is widely recognized that one cannot intend what one believes oneself to be incapable of doing (cf. Baier 1970: 658), and it seems clear that one cannot perform the actions of others (even though one can, of course, act on their behalf). Thus it seems to be impossible to include the actions of others in one's own intentions in the way it would be required in order to form intentions of the form "I intend that we J". Upon closer consideration, however, it seems that in these cases, one does not have to intend the actions of others in a straightforward sense, but that one simply has to take one's own intending to be of sufficient *influence* on the other participants so as to bring about their respective intentions to perform their part (Bratman 1999: 116).

element of collectiveness that is presupposed at its very base, because he endorses formal individualism.

In this respect, it seems to be Searle rather than Bratman who gets things right. For Searle stoutly opposes formal individualism. In his view, collective intentionality is a “primitive phenomenon” which is not to be reduced to any set of individual “I intends” plus mutual knowledge.²³ Yet Searle, too, sees himself confronted with the group mind, and he, too, resorts to individualism in order to banish it. However, in his conception, it is *subjective individualism* that plays that latter role. Searle argues that methodological solipsism is the only way to navigate safely between the two unacceptable alternatives, i.e. the Scylla of reductive formal individualism on the one hand and the Charybdis of the group mind on the other (the latter Searle calls “a perfectly dreadful metaphysical excrescence”²⁴). Thus he claims, in a modern version of Descartes’ *genius malignus* argument, that our collective intentionality is entirely in the heads of individuals and structurally independent of anything beyond individual minds.²⁵ Even a solitary brain in a vat that is somehow fed with the appropriate stimulus, or just lost in its dreams, and that is thus deluded about its real circumstances, could have intentions of the form “we intend”. In Searle’s view, the “we intend” (which is not reducible to individual “I intends”) is something single individuals have in their minds, and this is structurally independent of whether or not these minds stand in actual relations to the world – or to each other, for that matter.

Many of Searle’s critics think that this is wrong²⁶ – with good reason, I believe. It is true, of course, that the actual nonexistence of a group or the inexistence of co-members does not necessarily prevent individuals from intending *as if* they were members of that group. Just imagine the case of a dream about being one of the

²³ Among others, Searle puts forward the following two arguments against reductionism. First, common knowledge does not amount to the “sense of collectivity” involved in collective intending (1990). Second, our mind is too limited for the infinite iterations of knowledge implied in the “common knowledge” approach: “I think my poor brain will not carry that many beliefs” (Searle 1998b: 15).

²⁴ Searle (1998b: 150); see also Searle (1990: 404); Searle (1998a: 118).

²⁵ “Anything we say about collective intentionality must meet the following conditions of adequacy:

Constraint 1

It 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.

Constraint 2

It must 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. And this constraint applies as much to collective intentionality as it does to individual intentionality. One way to put this constraint is to say that the account must be consistent with the fact that all intentionality, whether collective or individual, could be had by a brain in a vat or by a set of brains in vats”; Searle (1990: 406ff.).

²⁶ Most forcefully, Anthonie W. M. Meijers has argued against the endorsement of methodological solipsism in the theory of collective intentionality; cf. Meijers (1994, 2002, 2003). See also Johansson (2003); Hornsby (1997); Waldenfels (1996); Celano (1999: esp. p. 239ff.); Turner (1999: 216, fn. 20).

dancers in the group in the first version of Henri Matisse's 'Dance'.²⁷ If we neglect the question about whether or not metaphysical socialism about intentionality is true, it seems obvious that minds that do not stand in actual relations to others, or brains in vats for that matter, may well *take themselves* to be members of a team. The decisive question, however, is whether or not even philosophers who accept subjective individualism concerning *individual* intentionality should be social externalists concerning *collective* intentionality (as I suggest they should): should we take such intentionality to be *collective intentionality* that just happens to be *mistaken* in some way, or shouldn't we rather say that this intentionality does not qualify as collective intentionality in the first place?

Searle advocates the first alternative. In his view, "the existence of collective intentionality does not imply the existence of collectives actually satisfying the content of that intentionality" (1997a: 450). For Searle, such cases as the one just mentioned simply show that "my presupposition that my intentionality is collective may be mistaken" (1990: 407). He admits that the case of a solitary brain in a vat having we-intentions constitutes a mistake of a very special kind²⁸ "which violates the Cartesian assumption that we cannot be mistaken about our intentions" (1998b: 150). But this "price to pay" (ibid.) seems all the more moderate since, in Searle's view, the Cartesian idea about the transparency of our intentionality proves to be wrong even in the case of individual intentionality, and is thus a mistaken notion anyway. Contrary to what Descartes thought, we can be mistaken about one of our intentional states²⁹ – why should this not be true for collective intentionality? In other words, the fact that there might be no actual collectivity involved in our collective intentionality boils down to just another way in which intentions can be mistaken – something that fails to touch the very *structure of our intentionality* itself.

Together with Searle's critics, I would like to put forward a different view. It seems to me that, by conceptually restricting collective intentionality to what is in individual minds, Searle misses a crucial element in the makeup of collective intentionality, which is the very element that Bratman emphasizes in his departure from solipsism and his move towards an interrelationalistic account of collective intentionality. For the sake of the argument, let's accept the *general* possibility of envatted brains³⁰ in order to take a closer look on Searle's claim that collective intentionality

²⁷ Cf. the reproduction on the cover of Searle's "Construction of Social Reality" (1995).

²⁸ What is in question here "is not simply a failure to achieve the conditions of satisfaction of an intentional state and is not simply a breakdown in the background"; cf. Searle (1990: 407).

²⁹ Cf. Searle (1998a: 69ff.). Here, Searle distinguishes four ways in which we can be mistaken about our consciousness in general, and our intentions in particular including self-deception (e.g. in the case of suppression of our dark sides) and misinterpretation (as in the case of somebody who takes his temporary infatuation to be real love).

³⁰ The possibility of 'envatted brains' is highly controversial. Putnam (1981) argues that it can be ruled out a priori; Dennett (1991) argues that the computational performance required in order to provide the 'envatted brain' with the appropriate input would be "computationally intractable on even the fastest computer". The question to be addressed here, however, is not whether or not 'envatted brains' are possible at all, but whether or not those brains, *if they were possible*, could be said to share intentions.

“could be had by a brain in a vat or by a set of brains in vats” (1990: 407). Imagine Ann and Beth visiting the Museum of Modern Art together; they happen to be the only visitors at the time. On the first floor, they get lost in the sight of the first version of Henri Matisse’s ‘Dance’ (the example is a homage to Searle’s [1995] cover illustration). Now a figure that is hard to avoid in envatted brains thought experiments puts in his appearance: an evil scientist creeps up behind our two heroines, and while Beth runs away screaming for help, he anesthetizes Ann for a minute, puts her brain in a vat and connects it to a computer that provides it with the appropriate input so that Ann has the impression of simply continuing to contemplate Matisse’s ‘Dance’ together with Beth, just as if nothing had happened. Now it seems that, in her vat, Ann still has intentionality that conforms to Searle’s concept of collective intentionality. All the intentionality Ann has in her mind seems to remain unchanged in subject, intentional mode, and content. It is still Ann’s intentionality, and she still intends to contemplate Matisse’s ‘Dance’ together with Beth (or, for that matter: she still intends her contemplating Matisse’s ‘Dance’ as her ‘we-derivative’ individual contribution to her and Beth’s shared intentional activity). Thus Ann may still have intentionality that is collective in *form* and that has ‘collectivity’ or ‘sharedness’ in its *content*. However, it is clear from the semantics of the verb “to share” alone that, in her vat, whatever she might *believe* she intends, Ann does not *in fact* share the intention to contemplate Matisse’s ‘Dance’ together with Beth anymore. It is obvious (and trivially true) that the sharedness of intentionality is not a matter of the *form* or *content* of one single individual’s intentionality alone. The question that turns out to be non-trivial is: what is it that has to be *added* to the picture for there to be proper shared intentionality?

In spite of its obvious importance to the theory of collective intentionality, Searle seems to be strangely disinterested in this question.³¹ It seems clear, however, that within his internalist framework, the following answer imposes itself: When Ann and Beth were in fact *sharing* their intention to contemplate Matisse’s dance together, they *both* (we-) intended to contemplate Matisse’s dance (or to contemplate Matisse’s ‘Dance’ individually as their contribution to their shared intentional activity). After the evil scientist’s intervention, however, *only Ann* (we-) intends to contemplate Matisse’s ‘Dance’ together with Beth. Beth, on her part, has no such intentionality any more, for she now intends to do something quite different, i.e. to run to the information desk of the Museum of Modern Art as quickly as she can to call for help. Thus it might seem that the answer to the question of what the intentionality Ann has in her vat lacks in order to qualify as *shared* intentionality can be found in Beth’s head. In order for (we-) intentionality to be shared, *all* participants have to have the appropriate (we-) intentions, which is not the case anymore in the given situation.

This answer, however, is clearly deficient. Here is why. Imagine the story of Ann, Beth, and the evil scientist to continue as follows. After the evil scientist has finished his business with Ann, he goes after Beth. On the ground floor, halfway to the exit, he catches up with her, anesthetizes her and puts her brain in a separate vat,

³¹ This has not escaped Bratman’s notice (1999: 116, 145).

connecting it to a second computer. Beth forgets all that has happened since the evil scientist appeared on the scene, and she is provided with the appropriate input so that the intentionality she has is "We contemplate Matisse's 'Dance'" or "I contemplate Matisse's 'Dance' as my part of our contemplating". Now let's get back to Ann, who is still in her vat on the upper floor. According to the internalist-minded view of the sharedness of intentionality I just sketched above, it seems that Ann's intentionality has become *shared intentionality* again in the very moment when the evil scientist switched on Beth's computer. For now, just as before the evil scientist's intervention, *both* Ann and Beth have intentionality of the form "we are contemplating Matisse's 'Dance' together" or "I am contemplating Matisse's 'Dance' as my part of our shared contemplating". This conclusion, however, is implausible; intentionality does not become shared intentionality just because completely independently of each other, two brains just happen to have appropriately 'matching' illusions. If shared intentionality is not a matter of what goes on inside an individual head *alone*, it is not a matter of what goes on inside *different* heads, either. In order to find out about the sharedness of Ann's and Beth's intentionality, it is not enough to check only what is in the minds of the two individuals. As Anthonie W. M. Meijers has pointed out most forcefully, sharedness is a matter of the *relations* between minds, i.e. something that "transcend[s] the boundaries of [. . .] the 'brain in a vat'" (Meijers (1994: 7).

The further question is: what *kind* of relation is required for intentionality to be shared? What *sort* of 'connection' do we have to add to the Searlean picture of isolated minds for there to be proper sharedness? I cannot aspire to giving a straightforward answer here, but shall restrict myself to contrasting my ideas with Meijers's, whose critical discussion of Searle's account of collective intentionality I still believe to be the most important one in the existing literature.

Meijers opposes Searle's theory of collective intentionality in at least two ways. Firstly, he argues that Searle's internalism has to be given up in favor of a relational account. Secondly, Meijers criticizes Searle's view that collective intentionality does not involve social normativity in the form of commitments, obligations, and entitlements.³² Along this line, Meijers argues that we have to give up Searle's *cognitivism* in favor of a *normativist* stance.³³ It seems that on Meijers's view, these two moves are internally connected, or even just two different aspects of one and the same move, so that the "radical relational approach" to collective intentionality he advocates somehow *has* to be a normativist one. This becomes clear from passages such as the following: "Cognitive attitudes are not sufficient to explain the *sharing* of intentionality. Normative attitudes have to be part of the analysis." Is Meijers right? And if not: why does he think the relations in question have to be normative ones?

In Meijers's view, collective intentionality "arises [. . .] out of the act of agreeing",³⁴ and it is within an analysis of this aspect of collective intentionality that

³² In Searle's view, any such normative phenomena come into play only with the use of language, which is logically posterior to collective intentionality (see Searle (2001a: Chaps. 5 and 6)).

³³ For a detailed normativist account of shared intentionality see Gilbert (1996).

³⁴ Meijers (1994: 89); cf. *ibid.*: 104ff., 143.

we have to go beyond Searle's internalism and move towards a relational account (Meijers 2003: 176, 167). Applying this view to the above example, it is essential for the very sharedness of Ann and Beth's intention to contemplate Matisse's 'Dance' together that there is some kind of (implicit) *agreement* between them, some shared *commitment* to do so, which to some degree *obliges* Ann and Beth to do their part and at the same time *entitles* both of them to rebuke the other if she does not perform her part.³⁵ Meijers argues that Searle's internalist theory of collective intentionality cannot account for these normative aspects. It seems clear that in her vat, Ann still might *believe* there to be an (implicit) agreement between herself and Beth to contemplate the paintings on exhibit together; however, as Meijers points out, there is a difference between *agreeing* and *seeming to agree* (Meijers 2003: 179), and it is this difference that the Searlean approach to collective intentionality cannot account for because of its internalist limitations. Just looking at what goes on in the individual mind of Ann there is no telling whether she is in an *actual* agreement with Beth or just *believes* herself to be so. In the latter case, however, there is no agreement and thus no shared intentionality between Ann and Beth.

I believe that this argument is sound in itself, but I do not see why the difference it hinges on – the difference between “A-ing” and “seeming to A” – should be specific to the normative aspects or forms of shared intentionality. It seems to me that the same point can be made within a cognitivist view, according to which the relations involved in collective intentional states do not necessarily have to be of the normative sort. Consider the following case. Imagine Ann and Beth to be dyed-in-the-wool Searleans. For them, their visit to the museum does not involve any kind of commitment, obligation, or entitlement whatsoever. However strange this might seem, it just happens to conform to their usual practice that any of them may walk away from the common enterprise at any time, without owing the other any further explanation. There is no agreement whatsoever between them; they are both just regular visitors to the museum on Sunday afternoons who over time have come to see their individual visits as part of a common enterprise. The first to come usually waits at the entrance for the other; if, as it sometimes happens, the other does not show up, she does not feel that the missing party has wronged her, or that she is entitled to some explanation. On the face of it at least, the intentionality involved is thus strictly limited to cognitive aspects. My point is the following: Even though there is no agreement, entitlement or obligation around, it still makes a difference if Ann just *believes she shares* the intention to visit the museum together or if she *actually shares* this intention. If Searle cannot account for the normative aspects of shared intentionality within his internalist framework, he cannot account for the purely cognitive aspects either.³⁶ Thus it does not seem necessary to connect the two

³⁵ For a normativist account of shared intentionality see Gilbert (1996).

³⁶ Indeed it seems that there is a great deal of purely cognitive components involved in shared intentionality. Take the case of Anne and Beth in their respective vats. What sort of connection has to be established between them in order for them to share intentions? It seems that a great deal of delusion is compatible with shared intentionality. Indeed there is even a sense in which the two brains in vats might said to be share their intentions, if their respective sources of input

issues Meijers raises against Searle. One does not *have* to take a normativist stance on collective intentionality in order to follow Meijers's advice to give up Searle's internalism in favor of a "radical relational approach" (Meijers 2003: 167).

§8 Collective Intentionality: Irreducible and Relational

The comparison between Bratman's and Searle's account reveals complementary strengths and weaknesses. On the one hand, Searle is right in renouncing formal individualism which seems to be the weakness of Bratman's account. Shared intentionality is not reducible to sets of I-intentions, because the I-intentions individuals *have* when taking part in a shared activity *presuppose* shared intentionality. On the other hand, Bratman is right in departing from subjective individualism. As Bratman makes clear (and contrary to what seems to be a consequence of Searle's approach), it is only *in relations* that individuals share intentions. Thus it seems that Bratman and Searle, in their respective departures from the Cartesian model of intentionality, both get stuck half-way in the project of developing an adequate account of collective intentionality. Their problem is that they let go of only one version of individualism, while holding on to the other. An adequate account of collective intentionality, however, has to depart from the Cartesian individualistic picture of intentionality not just in renouncing *either* formal *or* subjective individualism. It has to be *both* non-reductive *and* relational.

By way of giving a rough outline of my idea of such an account, I should like to propose two tentative theses, concerning the relational (i) and irreducible (ii) character of collective intentionality, respectively.

- i. Social normativity arises out of shared intentionality (and not the other way around)

Agreement-based accounts of shared intentionality beg the question because any sort of agreement *presupposes* shared intentionality. The act of agreeing is itself a move *within* a shared intentional activity (whereas not all cases of shared intentional activity involve agreement). Shared intentions which are based on agreement do, of course, exist. But these are shared intentions of a special (and especially complex) kind. Therefore, it seems that they should not be taken to be the 'paradigm case' of an analysis of shared intentionality. Thus I agree with Searle (as well as with Raimo and Maj Tuomela; cf. 2003b) that collective intentions do not *by themselves* involve social normativity in the form of obligations and entitlements. At the same time, however, I find the Searlean picture of completely normativity-free collective intentional activities (such as the one depicted above) rather askew. If the sharedness of intentionality is not necessarily *in itself* socially normative, it has *socially normative*

are appropriately connected (so as to make Anne believe that Beth does what Beth believes she does, and conversely). "The Matrix" provides a vivid illustration of shared cooperative activity by appropriately interconnected "brains in vats".

consequences. As was pointed out repeatedly in sociological theory, proper social norms arise out of merely habitual social practices such as customs (Geiger 1987). It seems to be almost inconceivable that we might engage in shared intentional activities over an extended period of time without our *cognitive* expectations concerning the actions of others gradually turning into *normative* expectations (which entails no less than a reversal of the direction of fit of the respective attitude; I shall come back to this in the concluding chapter of this book). My conjecture is that these socially normative *consequences* of shared intentions stem from the pre-socially normative (or, in Tuomela's words, from an "instrumentally 'normative'") *implication* of any kind of shared intention. The implication in question is the following. To the individuals involved, a shared intention provides a *reason* to form an appropriate personal intention (i.e. the intention to perform one's part).³⁷ In a pre-socially normative sense, if we intend, I *ought* to do my part in what we intend. This *normative* relation between shared intention and individual we-intention, however, does not exclude the possibility of overriding contrary reasons or simple weakness of will. Thus it seems possible that *we intend x* without me *intending to do my individual part* (even without my having a pro-attitude towards our shared aim; cf. Chapter 3 below). If this perspective on the relation between shared intentions and personal 'contributive intentions' is correct, it has far-reaching consequences: the analysis of shared intentionality cannot be based on an analysis of what individuals personally intend when sharing an intention, but the analysis of what individuals intend when taking part in a shared intentional activity has to be based on an analysis of the structure of shared intentionality (Tuomela 2002b). Or, to use Edmund Husserl's concept of *foundation* (Rota 1989):

- ii. Shared intentionality is the foundation of individual (we-) intentionality (and not the other way around)

The German phenomenologist Gerda Walther, thinking about shared intentionality in the early 1920s of the last century, and struggling against her own individualistic preconceptions, which she had taken over from Husserlian phenomenology, used a striking metaphor for what seems to be at stake in the move towards a non-reductive and relational account. She talked of a "Copernican Turn" (Walther 1923: 98) from an analysis of sharedness that is derived from an analysis of the intentionality of the participating individuals to an understanding of the intentionality of the participating individuals that is based on a solid concept of sharedness. But how should

³⁷ This seems to be at odds with John Broome's (2001) claim that intentions are not, as such, reasons for action. Broome's point is that if I have no justifying reason to intend to A, but intend to A (perhaps because I am mistaken about the relevant facts), I would have reason to do what I have no reason to intend to do, which does not seem plausible. In the current case, however, the issue at stake is not the relation between intention and action, but between collective and individual (participatory) intention. And as far as contributive action is concerned, it seems obvious that in normatively stabilized cases of joint action there are normative expectations from the other participant's side involved. Even if there is no justifying reason for us to intend J, I might have a reason to do my part; insofar as under some description, our aim is to do what we have reason to do, however, my part may well be to voice my doubts about there being a reason. For a more detailed discussion cf. Chapter 3 below.

such a turn be possible without simply replacing the individual with the collective as the source and bearer of intentionality? As seen above, fear of the group mind plays an important role in driving some of the most important accounts of shared intentionality back into the seemingly safe harbor of individualism. Thus it seems important to address the question: is this fear justified? Will an account that neither embraces formal individualism nor subjective individualism end up getting stuck with the group mind?

I believe that any such reservations against a non-individualist (i.e. non-reductive and relational) account of collective intentionality are mistaken. As seen, the whole trouble with the group mind arises from the attempt to give some acceptable answer to the question: who is the subject that *has* collective intentions? To whom can this intentionality be attributed as its source, bearer, or owner? And this question, innocent as it might look, is heavily loaded with historical ballast that we should, I think, simply jettison and leave behind.

Only in the last decades, have we successfully managed to get rid of Descartes' quest for absolute certainty in philosophy.³⁸ However, the Cartesian preoccupation with the "subject" still persists. It is still a deeply rooted idea that where there is intentionality there has to be a somebody who "has" it as its owner, source, or bearer.³⁹ It is the fact that most philosophers of collective intentionality hold on to this assumption that gives rise to the fear that by moving too far away from individualism, we are running the risk of getting stuck with the group mind. Yet there is a simple way out of the individualistic dilemma – or double-bind – in which current collective intentionality analysis seems to have gotten stuck: it consists in overcoming the "Cartesian Brainwash" by ceasing to address the "who has it" question. Collective intentions are not intentions of the kind anybody *has* – not single individuals, and not some super-agent. For collective intentionality is not subjective. It is relational. Collective intentionality is an intentionality which people *share*.

³⁸ Even Searle, who is by some accused of sticking to the Cartesian "epistemological" paradigm in philosophy (cf., e.g., Dreyfus 1993), says explicitly that he is not "a part of the Cartesian tradition of trying to overcome skepticism and provide a secure foundation for knowledge"; Searle (2001b: 173).

³⁹ The preoccupation with the subject or "bearer" of intentionality seems to stem from what is perhaps Descartes' most durable insight. I myself have a privileged position among all the things I might be acquainted with. However deluded I am about the world – and, we can add, about my intentions – there seems to be something incorrigible or infallible involved in my self-awareness. Even if I live in complete delusion about all my beliefs, there is still something that I simply cannot get wrong: it is in fact *myself* whom I am aware of when reflecting on my beliefs and desires. Even if some madness has me in its tightest grip, misleading me into thinking that I am Henri Matisse, it is still infallibly *me myself* whom I take to be Henri Matisse – it is not, for example, the actual Henri Matisse whom I take to be Henri Matisse. This insight seems to be at the base of Descartes' claim that what is really certain and indubitable about my thinking is the subject, the bearer of intentionality, i.e. the thinking "I". Now it seems obvious that, however right this might be concerning the "I" of individual intentions, it does not apply to the "we" of collective intentions, for I might easily be mistaken in any collective belief or intention.

In conclusion, I should get back to the initial ontological question concerning the relation between the individual and the collective. I think that with the illusion of the group mind the urge to drive actual collectivity out of the concept of collective intentionality vanishes, too. A theory of collective intentionality that is both non-reductive *and* relational does not require any logical or ontological primacy for the aims, attitudes and emotions of the individuals over the actual existence of the group. This does not mean, however, and conversely, that it requires the logical or ontological primacy of the group over the individual. Searle seems to think that we have to make our choice between these two versions of the Philosophy of Society: either we put the we-intentionality of individuals or the collective itself first (qua “ontological primitive” [Searle 1997a: 449] that somehow precedes our we-intending). His choice, then, is the first alternative: “Collective intentionality gives rise to the collective and not the other way around” (ibid.). It seems to me that, conceived of like this, the whole question about the relation between the individual and society is wrongly put. It implies what I should like to contest: that collective intentionality and actual collectivity are two different things. Only because, in the current debate, collectivity was driven out of the concept of collective intentionality in the first place does the question about how one is related to the other arise. If collective intentionality is not subjective, but relational, there is no need to postulate any ontological order of hierarchy between the analysis of collective intentionality and the ontology of groups. Because, in a relational sense, collective intentionality is what the ontology of groups is all about.

What is the bearing of this result on the Philosophy of Society? It seems that overcoming the ‘Cartesian Brainwash’ means to break away from the individualistic approach to Philosophy of Society, and to move towards a more heterodox view. In light of a post-Cartesian concept of collective intentionality, it appears that the orthodox slogan that “there is no society, only individuals who interact with each other”⁴⁰ is not outright wrong, but simply meaningless. Most forms of interaction involve collective intentionality, and collective intentionality is what society in the most basic meaning of the word *is*.

⁴⁰ Cf. Elster (1989: 259) quoting Margaret Thatcher. It should not be forgotten, however, that Thatcher continued as follows: “there are only individuals, *and there are families*” (Woman’s Own Magazine, 10/3/1987).